National and International University Departmental Web Site Interlinking: A Webometric Analysis

Xuemei Li, MSc

A thesis submitted in partial fulfilment of the requirements of the University of Wolverhampton for the degree of Doctor of Philosophy

December 2005

This work or any part thereof has not previously been presented in any form to the University or to any other body whether for the purposes of assessment, publication or for any other purpose (unless previously indicated). Save for any express acknowledgements, references and/or bibliographies cited in the work, I confirm that the intellectual content of the work is the result of my own efforts and of no other person.

Signature Xuemei Li

Date 15 Dec. 2005

Acknowledgements

First of all, I want to thank my first supervisor Prof. Mike Thelwall, a leading person in webometrics. It has been a real privilege to have him as my supervisor. He has always been an encouragement, giving me enough room to build up my own thinking and confidence in my subject. Mike has always been at hand to see me through difficulties. I am so appreciative of his guidance and kind support during the past four years.

I also want to thank my second and third supervisors, Dr. David Wilkinson and Dr. Peter Musgrove for their kind help during my research. Also many thanks to other group members in the Statistical Cybermetrics Research Group.

Thanks are also due to Dr. Henry Small from ISI and Linda Butler from the Research Evaluation and Policy Project, Research School of Social Sciences in the Australian National University. With ISI's permission, Linda has supplied the citation data for Australian departments in the project.

I would also like to thank the School of Computing and Information Technology in University of Wolverhampton for granting me the scholarship for the PhD project, and for supplying the facilities needed, especially during the link data collection, when hundreds of PCs were needed. Many thanks also due to the technicians and office staff for their kind support.

Thanks are also due to Prof. Mike Jackson for his kind advice for my transfer report. Thanks also to Dr. Clément Arsennault from Université de Montréal for his kind help enabling me to category the target pages in Quebec correctly. I would also like to thank my office mate and good friend Rudy Prabowo, who has been such an invaluable help and support during the four years.

I especially want to thank Mrs Deen Lloyd (mother Deen) who accommodated me for 18 months in her cosy house. She has been so patient in helping me through the grammar of the thesis. Finally, I want to thank my parents for their patience and support during the long PhD journey.

Publication List

Journal Papers:

Li, X. (2003) A review of the development and application of the web impact factor, *Online Information Review*, 27(6), 407-417.

Li, X., Thelwall, M., Musgrove, P. & Wilkinson, D. (2003b) The relationship between the WIFs or inlinks of computer science departments in UK and their RAE ratings or research productivities in 2001, *Scientometrics*, 57(2), 239-255.

Li, X., Thelwall, M., Wilkinson, D. & Musgrove, P. (2005a) National and international university departmental web site interlinking, Part1: Validation of departmental link analysis, *Scientometrics*, 64(2), 151-185.

Li, X., Thelwall, M., Wilkinson, D. & Musgrove, P. (2005b) National and international university departmental web site interlinking, Part 2: Link patterns, *Scientometrics*, 64(2), 187-208.

Thelwall, M., Vaughan, L., Cothey, V., Li, X. & Smith, A. G. (2003) Which academic subjects have most online impact? A pilot study and a new classification process, *Online Information Review*, 27(5), 333-343.

Wilkinson, D., Thelwall, M. & Li, X. (2003) Exploiting hyperlinks to study academic web use, *Social Science Computer Review*, 21(3), 340-351.

Conference Oral Presentation:

Li, X., Thelwall, M., Musgrove, P. & Wilkinson, D. (2002) The relationship between the links/Web Impact Factors of computer science departments in UK and their RAE (Research Assessment Exercise) Ranking in 2001, *Seventh International S&T Indicators Conference*, Karlsruhe Germany, 25 to 28 September 2002.

Li, X., Thelwall, M., Musgrove, P. & Wilkinson, D. (2003a) Do academic web pages in Western Europe attract more Links if they are in English?, *9th ISSI*, Beijing, China, 25-29 August 2003.

Abstract

In recent years, the structural similarity between hyperlinks and citations has encouraged information scientists to apply bibliometric techniques to the Web, with the hypothesis that studies of links may reflect patterns of informal scholarly communication in the way that citations can be used to illustrate formal scholarly communication. University web site interlinking has consequently been extensively investigated, but much less is known about departmental interlinking, i.e. links to, from, or between sets of departmental web sites. University web sites are large compared with departmental web sites, and statistically significant results are more easily obtained. Nevertheless, universities are multidisciplinary by nature and various disciplines may employ the Web differently, thus patterns identified at the university level may hide subject differences. Departments are typically subject orientated, and departmental interlinking may therefore illustrate interesting disciplinary linking patterns, perhaps relating to informal scholarly communication. Similarly, international academic interlinking at the departmental level is another relatively neglected research area.

The research aim of this thesis is: firstly to validate departmental link data; secondly to identify whether and how link patterns differ along country and disciplinary lines amongst *similar* disciplines and *similar* countries. In order to do so, physics, chemistry and biology departments in Australia, Canada and UK were chosen. The subjects are all hard sciences, and are therefore relatively similar, and potentially able to reveal subtle differences in linking patterns. The three countries are all economically advanced, and all are predominantly English speaking, except for Quebec, which is a French speaking zone of Canada.

Techniques originally designed mainly for university link analysis are applied in this study, although modified to cope with the additional difficulties of an international department-based investigation. Both the commercial search engine AltaVista and the personal web crawler SocSciBot are used to collect the necessary link data. Significant correlations between inlinks and research quantitative indicators are present, and also between Web Impact Factors (with academic staff members as denominators) and research averages. The statistically significant results together with the results of a target page classification exercise serve to support the validity of the departmental link analysis as reflecting academic activities.

Citation counts are the most relevant data to compare with link counts, since the similarity between the two triggered webometric studies. Citation counts for Australian departments from 1998 to 2002 are from Research Evaluation and Policy Project (REPP), in the Australian National University. With regard to citation counts for UK and Canadian departments, the thesis introduces a technique to count citations in a semi-automatic way. In addition to citation counts, the results from the Research Assessment Exercise (RAE) 2001 are employed for the UK departments' correlation tests, while research grants received in 2003 from the Canadian National Sciences and Engineering Research Council (NSERC) are used for Canadian departments.

In order to get a holistic picture of departments' web use profiles and link patterns, four different aspects are identified for each set of departments. The four aspects are:

- General web use
- National peer interlinking
- International peer interlinking
- Interactions with different top level domains

Different link patterns are identified along both national and disciplinary lines. Along national lines, a likely explanation for the difference is that countries with better research performances make more general use of the Web; and, with respect to international peer interlinking, countries that share more scholarly communication tend to interlink more with each other. Along disciplinary lines, it seems that departments from disciplines which are more willing to distribute their research outputs tend to make more general use of the Web, and also interlink more with their national and international peers.

In summary, the country and disciplinary link patterns identified are both influenced by offline factors that affect the relationship between the source and target page owners. There can be significant differences in the way that similar disciplines use the Web, and these can point to underlying differences in informal scholarly communication.

Table of Contents

1. General Introduction

1.1 Introduction

The World Wide Web is unorganized and anarchic by nature (Winship, 1995). However, it has attracted interests from scientists in different subject areas, such as computer science, theoretical physics, information science, communication studies and sociology (Thelwall, 2004b). A belief is that useful information can be derived from it, especially through hyperlinks, which connect web pages together. This has been strengthened by Google's successful use of a link-based algorithm to enhance its performance (Brin & Page, 1998; Google, 2002).

The structural similarity between hyperlinks and citations has inspired information scientists to apply bibliometric techniques to the Web (Thelwall, 2004b). This has started a new research area: webometrics (Almind & Ingwersen, 1997). Chapter 2 reviews the transference of related bibliometric techniques to webometric studies such as the following.

- The Web Impact Factor (WIF) is the application of the Journal Impact Factor (JIF) to the Web (Ingwersen, 1998).
- Link Propensity (LP) is based on Salton's measure of co-authorship (Smith & Thelwall, 2002).
- Link counts correlate significantly with research measures as citation counts do (Thelwall, 2001b).
- Mathematical regularities, which are similar to those identified in formal publications, have been identified on the Web (Rousseau, 1997; Broder et al., 2000; Price & Thelwall, 2005, to appear).
- Co-linked and co-linking, which are from co-citations and bibliographic coupling in bibliometrics, help to identify useful clusters on the Web (Larson, 1996; Thelwall & Wilkinson, 2004).

The above evidence gives some confidence that it is feasible to further apply bibliometric techniques to webometric studies. Most academic link analyses have operated at the university level and within a single country. This is because firstly, universities are large enough aggregation units for conducting significant statistical studies; secondly, international link analysis is more complicated. However, universities are multidisciplinary by nature, and various disciplines may employ the Web differently (Kling & McKim, 2000; Vaughan, 2002; Tang & Thelwall, 2003c). Link patterns identified at university level may hide these differences. Departments are typically subject orientated, and departmental link analysis has the potential to illustrate disciplinary link patterns. In this respect, this study concentrates on link analysis at departmental level and across country boundaries. This is also motivated by various citation patterns identified amongst different countries and disciplines in bibliometrics (Garfield, 1999; Glänzel & Schubert, 2001).

Although the analogy between hyperlinks and citations triggered webometric studies, they are from different environments (Björneborn & Ingwersen, 2001). Patterns that have been identified through citation analysis are typically regarded as formal scholarly communication, since journal articles form a formal portion of research outputs. In contrast, those identified through links are regarded as informal scholarly communication, as web pages contain various contents with very limited amount of formal research outputs (Wilkinson et al., 2003). As a result, caution must be exercised when applying bibliometric techniques to the Web. Most techniques from bibliometrics need to be adjusted before their application in webometrics. For example, the WIF has undergone extensive enhancements before it could be used reliably on the Web (Li, 2003), as discussed in section 2.5. There are also some techniques that are solely devised for webometric purposes. For example, the Alternative Document Model (ADM) is a technique used to count links through different heuristic-based models of web 'documents' (Thelwall, 2002b), as discussed in section 2.4.

Various techniques which have been applied at the university level, for example, the WIF, LP and Alternative Document Models, can also be applied at the departmental level. The next section describes briefly these terms. However, link analyses at departmental level are more complicated than those at university level. It is more difficult to collect both the web and non-web data for departments than for universities. For example, the number of departmental interlinks is sparse compared with those of universities (Harries et al., 2004), and there are less departmental reports with regard to authoritative research measures and number of academic staff members.

1.2 Key Terminologies

This section defines and explains some basic link terminologies that are frequently used in this thesis.

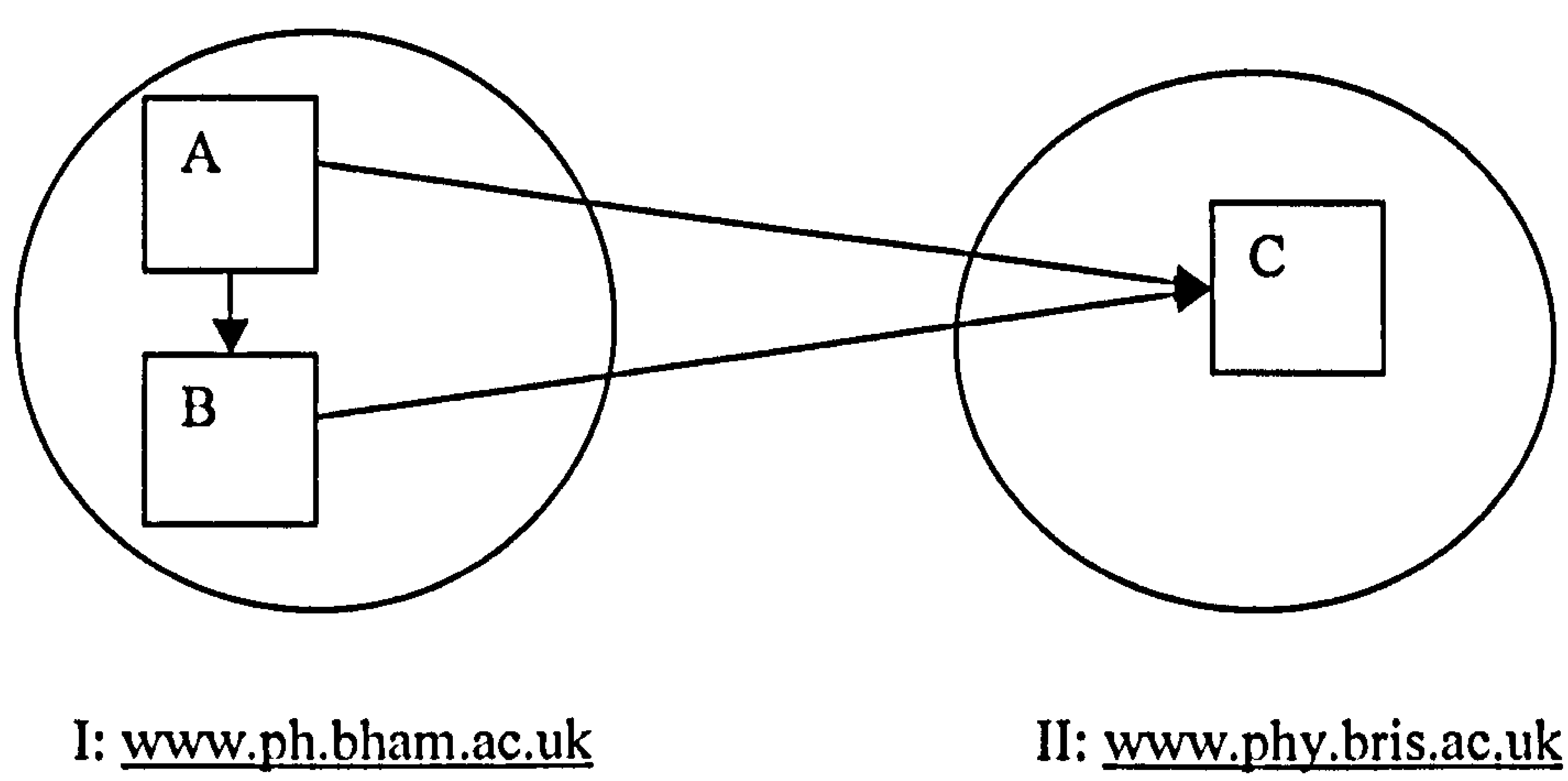


Figure 1.1 Links between two Physics departments

Figure 1.1, a simple example, is based on Björneborn's approach, which shows link relations between web nodes (Björneborn, 2001a). Circles stand for departments' web sites. Squares stand for web pages. Arrows between web pages, A, B and C represent links. The left circle, I: www.ph.bham.ac.uk, is the physics department in the University of Birmingham. The right one, II: www.phy.bris.ac.uk, is the physics department in Bristol University. The link from A to C is an external inlink to II. At the same time, it is also an external outlink from I. The link from A to B is an internal inlink (site self-link) of I. B and C are co-linked by A, while A and B are co-linking to (or coupled by) C.

Links from A and B to C are interlinks between the two physics departments. I has two web pages, while II has one web page.

The Web Impact Factor (WIF) is the number of inlinks received by a web site divided by the number of web pages within that web site (Ingwersen, 1998). The denominator can also be the number of academic staff members if the entity in question is a department or university (Thelwall, 2001b). Section 2.5 discusses the development and application of the WIF in detail. Another related link based metric is the Web Use Factor (WUF) (Thelwall, 2003b), which is the number of outlinks from a web site divided by the number of web pages within that web site. The only difference between the two is that the numerators are different.

Link Propensity (LP) is the number of external inlinks to a university web site or collection of such sites, divided by the product of the number of academic staff members from both the source and target universities (if numbers of staff members are not available, numbers of web pages are used instead) (Smith & Thelwall, 2002). It shows the tendency of two web sites linking to each other. Section 2.9.4 describes the LP in detail.

The Alternative Document Model (ADM) is a technique to count links between web nodes at different aggregations of document levels through manipulating their urls (Thelwall, 2002b). The rationale is to remove large numbers of repeated links, which may render other individually created links meaningless, at a higher document level. Section 2.4 discusses this technique in detail.

More terminology can be found in glossary of the thesis.

1.3 Research Aims and Objectives

This research aims at identifying departmental link patterns. In brief, the main objectives for this project are to

- Validate departmental link data
- Identify departmental link patterns along country and disciplinary lines

1.3.1 Validation of Departmental Link Data

In order to validate departmental link data, both correlation tests between research measures and link counts, and a target page classification scheme have been conducted in this study. This serves to test whether links to departments associate with research. If the result is positive, then the link patterns identified may disclose underlying informal scholarly communication on the Web amongst departments, which may be ignored otherwise.

1. Correlation tests

Citation counts are regarded as the most relevant data set to compare with link data, since the analogy between them triggered webometric research. Link counts have been compared with citation counts at three different levels. These are:

- Link versus citation counts
- Link counts per academic staff member versus citation counts per academic staff member
- Link counts per web page versus citation counts per publication (CPP)

A technique is introduced to count citations for departments in a semi-automatic way from the Web of Science (ISI, 2005). Section 5.2.4 describes this in detail. When other research quantitative indicators are reported at university level or do not exist at all, citation counts can always be a choice in the correlation tests.

In addition to citation counts, other publicly accessible research measures have also been applied in the correlation tests to further validate the departmental link data from different aspects.

2. Target page classification scheme

Since significant correlations do not imply causation, link motivation analysis must be conducted to further interpret link patterns identified. A target page classification scheme, which is completed by visiting source pages when necessary, is applied for this purpose. With regard to sparse numbers of departmental interlinks, whole population of target pages were studied. Section 5.4 describes this issue in detail.

1.3.2 Identification of Departmental Link Patterns

In order to get a full picture of the departmental link patterns, four different aspects can be studied. These are:

1. General web use
2. National peer interlinking
3. International peer interlinking
4. Interactions with different top level domains

The first aspect illustrates a set of department's web capacity with regard to their ability of to publish, to make and to attract links on the Web. The second shows how well a set of departments interlinks with their national peers. The third shows the ability of a set of departments to attract inlinks from international peers. The fourth has the potential to illustrate how a set of departments interacts with other web areas.

1.3.3 New Departmental Link Analysis Techniques and Methods Needed

This study concentrates on investigating new techniques and methods, which can be applied to

1. Collecting departmental link data. Existing techniques for university link data collection must be adjusted to suit their uses for departments. Section 5.1.2 describes these in detail.
2. Conducting correlation tests. The literature shows that authoritative research measures, which are used in the correlation tests, are mainly reported at the university level. A semi-automatic technique to count citations for departments is needed to overcome this problem. Section 5.2.4 describes this in detail.
3. Identifying various departmental link patterns. Four different aspects of departmental link patterns have been investigated in order to better understand the phenomena. A new link based metric is needed to indicate the ability of a set of departments attracting international peer inlinks from two other countries. Section 5.5 explains this in detail.

1.4 Research Contributions

This study proposes:

1. A method to collect departmental link data;
2. A semi-automatic technique to count citations for departments, in order to carry out the correlation tests;
3. A method to validate departmental link data;
4. A novel link based metric, which can be used to indicate a set of departments' ability in attracting international peer inlinks.

Finally, this thesis presents a method that can be used to identify departmental link patterns along country and disciplinary lines from four different aspects.

1.5 Thesis Outline

Figure 1.2 describes diagrammatically the outline of this thesis.

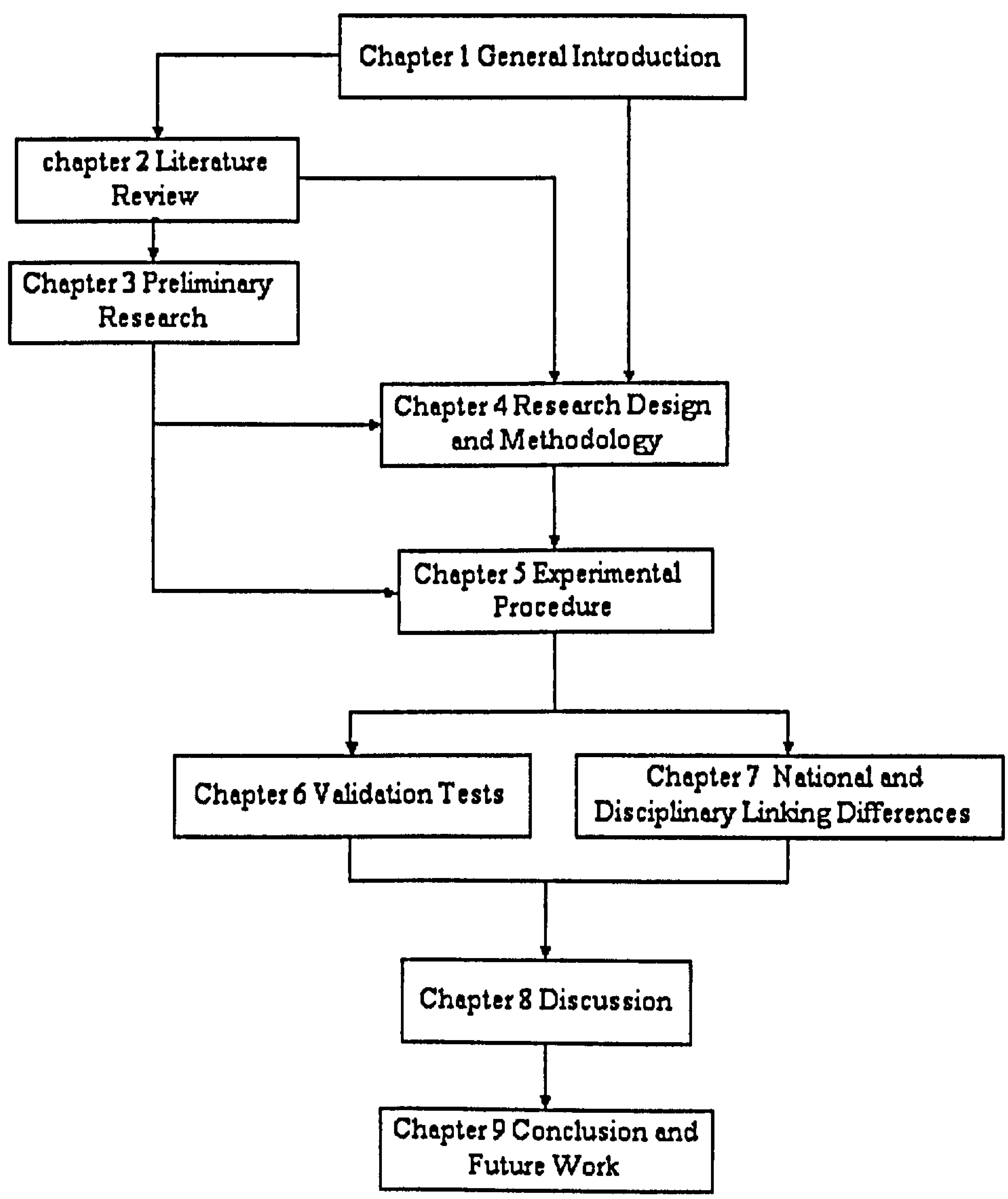


Figure 1.2 Thesis Outline

Chapter 2 reviews related work. The underlying connection between webometrics and bibliometrics is the analogy between hyperlinks and citations. The similarities and differences between citations and hyperlinks are discussed in section 2.2. Section 2.3 compares search engines and a bespoke web crawler SocSciBot with regard to their pros and cons in link data collection. The ADM technique is reviewed in detail in section 2.4. Section 2.5 reviews the origin and development of WIF. Section 2.6 discusses the correlation tests for link data validation in the context of bibliometric studies. Section 2.7 reviews various research measures for 'evaluating research output'. Section 2.7.1 broadly reviews existing research measures, while section 2.7.2 discusses and compares those that have been applied in link data correlation tests. Citation counts were regarded as the most relevant ones. Section 2.8 discusses link motivation analysis in the context of citation motivation analysis. Finally, section 2.9 discusses techniques applied in previous link analysis to identify link patterns.

Chapter 3 describes two preliminary findings before the main study in this thesis. They support the research design described in chapter 4, and the experimental procedure in chapter 5.

The first preliminary finding deals with the correlation tests between link and research measures for the 79 UK computer science departments who submitted in the RAE 2001. The significant result gave some confidence that departments are large enough units to carry out link analysis. The departments' link structure was extracted from existing universities' link data from SocSciBot without repeat crawling. Inlinks from different domains to the departments were counted through AltaVista.

The second second preliminary finding is that a higher proportion of English pages in a university web site attracts more international inlinks, at least amongst the fourteen Western European countries analysed. As a result of this, the departments in this study are chosen from English speaking countries: Australia, Canada and the UK. This is to ensure that link patterns identified are not affected by the linguistic effects.

Chapter 4 describes the research design and methodology that are applied in this study. A quantitative method is applied in this study. This chapter describes the research hypothesis, research design and methods. The main research hypotheses are: departmental link data is a valid information source to disclose informal scholarly communication between departments; and departmental link patterns differ along country and disciplinary lines, even for similar subjects. Section 4.1 explains the motivations for this research hypothesis. Section 4.2.1 discusses why the populations for this study have been selected. Section 4.2.2 describes the design structure of departmental link data validation, which includes correlation tests between research measures and link counts, and a target page type classification scheme. Section 4.2.3 discusses the design structure of link pattern identification, which is investigated through four different aspects. Section 4.3 describes the general research procedure and statistical methods used in this study.

Chapter 5 describes the experimental procedure details used to conduct the research, step by step. This follows the general research procedure introduced in section 4.3.1. Chapter 6 reports the results of the validation tests: correlation coefficient values and the departmental target page types. Chapter 7 reports the results with regard to the four aspects for each set of departments. Different link patterns were then identified along both country and disciplinary lines. Departments with outstandingly good or bad web exploitation were reported.

Chapter 8 discusses the difficulties encountered for departmental link analysis, with regard to web and non-web data collection, and the classification of the target page types. Apparent outliers in the linear diagrams of chapter 6 were identified, and reasons were tracked down. Link patterns identified in chapter 7 were also discussed to seek the underlying reasons. Chapter 9 concludes the project and suggests some interesting future research directions.

1.6 Summary

The structural similarity between hyperlinks and citations has encouraged information scientists to apply bibliometric techniques to webometric studies. Previous academic link analysis is mainly at university level, while relatively less has been studied at departmental level. Universities are multidisciplinary by nature, and link patterns identified at university level may hide different web uses from various disciplines. In this context, this study focuses on departmental link analysis, since departments are subject oriented and departmental interlinking may therefore illustrate interesting disciplinary link patterns.

In this respect, this study addresses the following two main issues:

- Whether departmental link counts are valid to convey useful information about informal scholarly communication.
- Whether there are any differences in link patterns for departments along country and disciplinary lines.

The first issue deals with both correlation tests and link motivation analysis. In contrast, the second issue investigates into four different aspects to illustrate a holistic picture of various departmental web uses along country and disciplinary lines.

This study reviews relevant academic link analysis in the context of bibiometric studies. Existing techniques and methods were discussed and compared to find out whether they are suitable for this study. Some techniques that have been used in link analysis at university level, such as, WIFs, LPs and ADMs, can also be applied at departmental level. However, departmental link analysis involves more complexities. This study concentrates on investigating new techniques and methods that are suitable for link analysis at departmental level.

The departmental link data has been collected through AltaVista and SocSciBot. This study introduces a method to extract departmental link structure from existing university link data from the SocSciBot without repeat crawling.

Citation counts are regarded as the most relevant research measures to carry out the correlation tests. This study proposes a semi-automatic technique to count citations for departments from the ISI's Web of Science.

With regard to link motivation analysis, the study proposes a classification scheme for target pages, which is completed by visiting source pages when necessary.

In addition to LP, 'adapted mean international peer inlinks' is proposed to indicate a set of departments' ability to attract international peer inlinks from two other countries.

In short, this study focuses on departmental link data validation and link pattern identification.

2. Literature Review

2.1 Introduction

The work presented in this thesis focuses on departmental level link analysis. Since the structural similarity between hyperlinks and citations has encouraged information scientists to apply bibliometric techniques to webometric studies, section 2.2 describes the similarities and differences between hyperlinks and citations.

In particular, this thesis describes how to:
(1) Validate a set of link data. A search engine and a bespoke web crawler are used to collect the link data. Section 2.3 describes each of them in detail. A webometrics technique, called the Alternative Document Models (ADM), has been used to count the number of links at different aggregated document levels. A link-based metric, called the Web Impact Factor (WIF) is applied to compute the normalised number of links between departments. Sections 2.4 and 2.5 subsequently describe each of them, and explain the reasons why they are applied in this study.

Given the number of links collected and a set of WIF values, two approaches are used to validate the link data. These are

(a) Correlation tests. They are used to calculate correlations between the link data and a number of research quantitative indicators. Section 2.6 describes this issue in detail. Section 2.7 discusses different research measures for the correlation tests.
(b) Link motivation analysis. Section 2.8 describes a number of different approaches which are used to identify the link motivation, and compares with those applied in citation motivation analysis.

(2) Use the link data to identify patterns. Section 2.9 describes the existing techniques for identifying patterns from link data in the context of bibliometrics.

2.2 The Similarities and Differences between Hyperlinks and Citations

Hyperlinks and citations are from two different environments, with citations in traditional print publications and hyperlinks in web publications. However, the structural similarity between the two (directional links between documents), has attracted information scientists to apply techniques from bibliometrics to the Web (Larson, 1996; Almind & Ingwersen, 1997; Rodríguez i Gairín, 1997; Ingwersen, 1998; Cronin, 2001; Prime-Claverie et al., 2002; Prime-Claverie & Beigbeder, 2004).

2.2.1 The Similarities between Hyperlinks and Citations

The similarities between hyperlinks and citations have been underscored by the new term 'sitation' introduced by McKiernam (1996) and Rousseau (1997). Cronin (2001) claimed that 'hypertext and citation indexing are a marriage made in heaven.' Link analysis is thus regarded as a nature growth from citation analysis (Cronin, 2001; Borgman & Furner, 2002).

The structural resemblance between the two can be identified clearly through the graph theoretic point of view. Regardless of the content, the Web may be viewed as a directed graph, where web pages are connected by hyperlinks (Kleinberg, 1999; Broder et al., 2000; Thelwall, 2001a; Meghabghab, 2002; Prime-Claverie & Beigbeder, 2004; Thelwall & Wilkinson, 2004). In bibliometrics, citation networks of scientific publications are created by papers connected with citations (Garner, 1967; Hummon & Doreian, 1989; Egghe & Rousseau, 1990; Small, 1999b). Björneborn (2004) uses a web node diagram to illustrate the link structure and define the basic link terminology. Based on his approach, figures 2.1 and 2.2 show the structural similarities between hyperlinks and citations. Table 2.1 lists the structural comparison between the two diagrams.

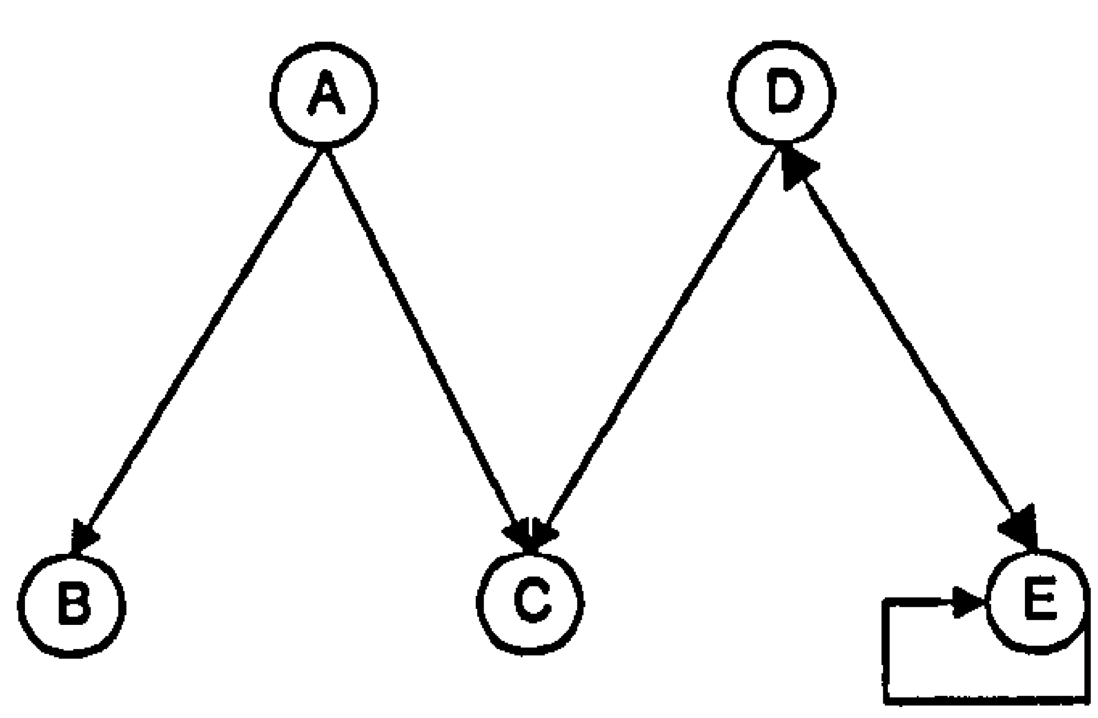


Figure 2.1 Web pages and hyperlinks

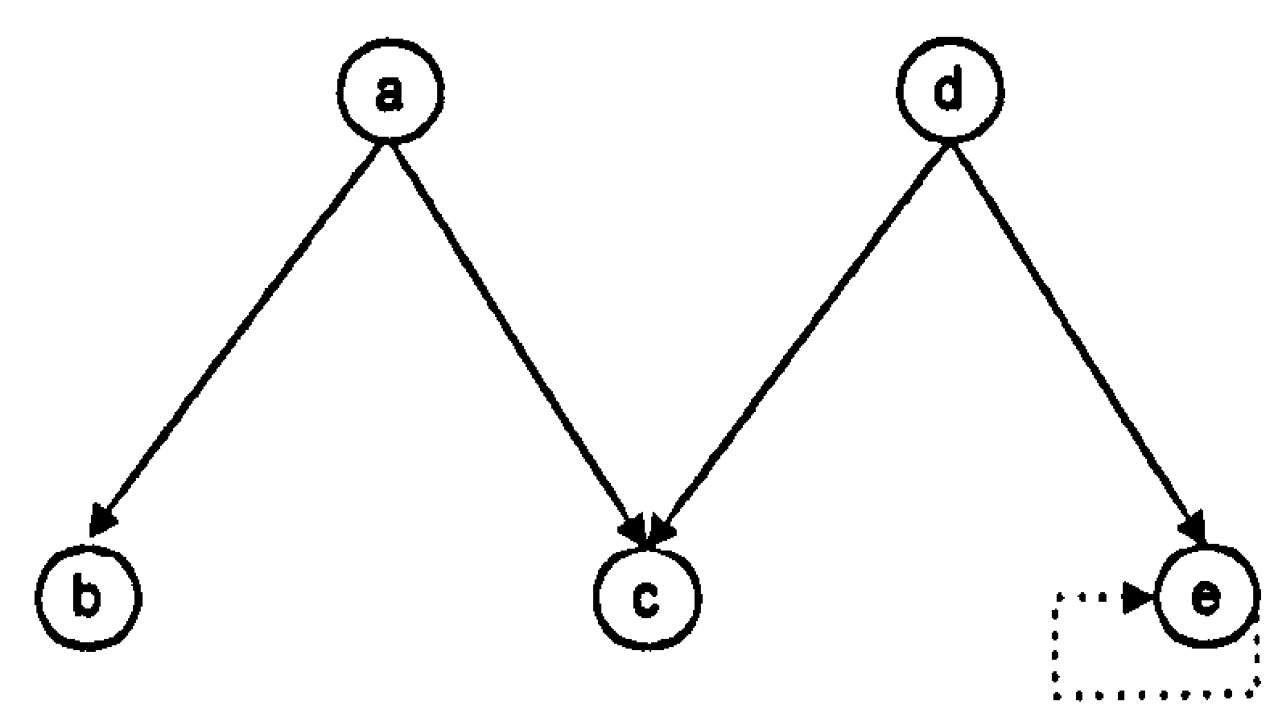


Figure 2.2 Papers and citations

The nodes in figure 2.1 can also be research units in webometrics. In addition to a web page, each node may represent a directory, a domain or a web site for a research group, a department, a university or a nation. In contrast, in figure 2.2, each node may represent a paper, an author, a journal, a research group, a department, a university or a nation. In figure 2.2, the self-citation line for node 'e' is a dotted one since a paper cannot cite itself, although the other units can do.

Table 2.1 Structural comparison between hyperlinks and citations

Figure 2.1	Figure 2.2
A has an outlink to B	a has a reference to b
B has an inlink from A	b has a citation from a
B and C co-linked by A	b and c co-cited by a
A and D co-linking to C	a and d coupled by c
E has a selflink	e has a self-citation (not for a node as a paper)
D and E link to each other at the same time	d and e cannot link to each other at the same time

2.2.2 Some Key Differences between Hyperlinks and Citations

Although hyperlinks and citations are similar in structure, they are different in nature. Citations are references given in books and articles, while hyperlinks are made between web pages. The dynamic, distributed, unorganised, anarchic nature of the Web implies that hyperlinks are more complicated than citations. Therefore, the analogy between the two should not be taken too far (Björneborn & Ingwersen, 2004). Some key differences are as follows:

1. Different quality control. Articles have to pass through strict peer review; normally at least two reviewers are required to confirm that the quality of a

paper is suitable for publication. In comparison, the content of web pages often lack any quality guarantee. Web creators have the freedom to publish anything or point to anywhere on the Web.

2. Different motivation types. Although there are more and more online journals, a serious issue is that any sort of content can appear on the Web. In addition to peer reviewed papers, lecture notes, preprints of draft papers, authors' CVs, hobbies, family information and religious content may appear on academic staff members' personal web pages (Wilkinson et al., 2003; Bar-Ilan, 2004a; Harries et al., 2004).
3. Different stability. Once citations between papers have been established, they are irreversible (Glänzel, 2003). In contrast, web pages and hyperlinks are dynamic since they can be updated easily (Almind & Ingwersen, 1997; Lawrence et al., 2001; Koehler, 2002; Markwell & Brooks, 2002).
4. Different certainty. Normally papers have one or many authors, and a clear publishing date. However, it is difficult to get the exact publication date of web pages, and also authors of web pages can not always be identified (Egghe, 2001).
5. Different currency. The information from the Web is normally more current than books and papers. Some new techniques, which cannot be found in books or papers, may be found on the Web.
6. Self-links are different from self-citations. A self-link is an inlink from the same site, which is also called an internal link (Ingwersen, 1998; Smith, 1999a). A self-citation is defined as a citation where the citing and cited papers share at least one author (Snyder & Bonzi, 1998; Borgman & Furner, 2002; van Raan & van Leeuwen, 2002). Self-links mostly serve navigational purposes rather than to endorse the target pages (Smith, 1999a; Björneborn & Ingwersen, 2004). Self-citations, however, usually indicate that authors build on their previous work (Bonzi & Snyder, 1991; Katz & Hicks, 1997; Glänzel & Thus, 2004), although with accusation of egotism (Baird & Oppenheim, 1994).
7. Structural difference. Citations are uni-directional, while hyperlinks can be bi-directional, as shown in table 2.1. Web pages can link to each other regardless of their publication date. In comparison, only earlier published papers can be cited by later published ones, not vice versa, although it is also possible for authors to cite each other's paper at the same time through 'invisible colleges' (Egghe, 2000). A web page can link to different parts of itself. However, a paper cannot cite itself.

The differences between hyperlinks and citations, however, have not sounded the death knell for webometrics, although they call for research to establish how hyperlink analysis relates to citation analysis. As part of this, WIFs have been found to correlate with non-web research measures not only at university level but also at departmental level (Thelwall, 2004b). The majority of links between universities are academic related (Wilkinson et al., 2003). Link patterns therefore illustrate informal scholarly communication, which ensure that 'the sometimes overlooked inputs and influences of technicians, mentors, trusted assessors and sundry collaborators could more easily be factored into the recognition' (Cronin, 2001). Other informal scholarly communication, such as those through private conversations, e-mails, meetings, and unpublished letters etc. may be lost without any record.

2.3 Web Data Collection

In order to quantitatively investigate the Web, reliable data must first be collected. Two search engines and a bespoke web crawler have been used for collecting academic link data. They both have advantages and disadvantages. Search engines are suitable for large web areas, while a bespoke web crawler is for small web areas.

2.3.1 Search Engines

Search engine performance is a very important component in webometric research, as earlier information scientists relied on search engines to collect link data. AltaVista (2001) and AllTheWeb (2002) have been used to count the number of link pages or web pages for webometric research (Ingwersen, 1998; Smith, 1999c; Chu et al., 2002; Smith & Thelwall, 2002; Vaughan, 2002), because they have relevant advanced search facilities (Sullivan, 2001a). Smith & Thelwall (2002) used both AltaVista and AllTheWeb in their study. The syntaxes (now defunct) below were used to count the number of link pages linking from the UK academic domain to Australian universities.

AltaVista: host:.ac.uk AND link:xxx.edu.au
AllTheWeb: url.host:ac.uk+link.all:xxx.edu.au

Where xxx stands for the third level of the domain name of an Australian university.

Google (2002) also has an advanced search facility, but it does not support the same level of Boolean querying as AltaVista or AllTheWeb. Google can only count all web pages linking to a given web page, but not to a given site. Its advanced search can limit the source to a given domain, but it cannot explicitly exclude all links from within the site itself, a second critical gap in its functionality for webometric data collection. Section 2.5.2 discusses the necessity of removing internal inlinks (site selflinks). Although Google is the most used search engine at the moment (Sullivan, 2001b; Sullivan, 2002), it is not recommended for collecting link data for link analysis purposes because of these limitations.

It is free and convenient to use search engines to collect link data. Especially for a large web area, search engines can be the only choice for collecting link data, because it is not pragmatic for a self-designed crawler to cover the whole Web or even the web area of a single nation.

Nevertheless, it is not reliable to collect link data through commercial search engines. The reason being:

1. Limited and uneven coverage of the Web. Given the large size of the Web, search engines can only index proportions of it (Lawrence & Giles, 1999). Furthermore, they cover the Web unevenly. Firstly, different search engines cover different parts of the Web, and the overlap among their coverage is very small (Bharat & Broder, 1998; Bar-Ilan, 2001). It is not surprising that different search engines return various results for same queries (Snyder & Rosenbaum, 1999). Secondly, one search engine not only may cover some domains more than the others, but also cover some web sites more than the

others within one domain (Thelwall, 2000). Search engines find new web pages according to: user registrations of new sites and links from previously indexed pages. The results from a search engine can only be relatively reliable, if the web sites searched are similarly registered and covered by the search engine.

2. Fluctuations of search results. Search engines may lose information. URLs that were returned by a search engine may disappear at a later query from the same search engine, even if the urls continue to exist and are relevant to the search topic (Bar-Ilan, 1999; Mettrop & Nieuwenhuysen, 2001). Results from same search engines may not be consistent even within one day (Bar-Ilan, 1999; Rousseau, 1999).
3. The opaqueness of search algorithms. Search engines are mostly for business purposes. Search algorithms are normally commercial secrets. They can be changed drastically without acknowledging the users (Pandia, 2004).

Hence, search engine results can only be regarded as indicative rather than definite values. In order to get reliable results, it is necessary for information scientists to create their own crawler (Bar-Ilan, 2001). The next section describes such a crawler in detail.

2.3.2 SocSciBot: A Bespoke Web Crawler for Academic Web Sites

Thelwall (2001f; 2001c) has designed a web crawler named SocSciBot to overcome some of the problems of search engines. Essentially, the crawler starts from the home page of a university web site, extracts all its links and then downloads all of the pages found that are on the same site. This process is repeated until all links have been followed. It more rigorously identifies and eliminates duplicate pages within a web site, as well as mirror sites that are not created by the staff members or students within a university. Through the use of the data collected in this way, researchers have more control over the extent of coverage of sites, and can also be in control of the algorithm used to count links from the database.

An important advantage to SocSciBot is its ability to deliver link data in such a form that it can be used in more complex counting methods. Alternative Document Models (ADMs) can be used to count the number of links at different document levels, for example, to count links among directories, domains or whole university sites. This will be discussed in detail in section 2.4.

A key drawback to the SocSciBot, as mentioned before, is that as a personal crawler, it is not suitable for large-scale studies. Another drawback is that personal crawlers can only cover the publicly indexable web pages: those that can be accessed by following links from a homepage, or a set of seed urls if the homepage does not contain any links. Web pages that are not linked directly or indirectly by the homepage or the seed urls of a university will be missed, even if they are linked to by web pages outside the university. Search engines may index the latter, as they identify web pages by following links from its known url lists, which are for the whole Web rather than a limited web site.

2.3.3. A Comparison between Search Engines and the SocSciBot

Search engines have to cover a significant proportion of the Web, and the priority of a search engine is to respond to a user's query efficiently rather than comprehensively. A search engine that is slow or does not tend to provide relevant results in the first page is likely to lose customers. Being business-driven, they may sacrifice their accuracies for performances (Bar-Ilan, 2001). Compared with search engines, the SocSciBot enables a more scientific approach by covering individual web sites comprehensively within specified parameters (Thelwall, 2002e). The counts of links or web pages from the databases collected by SocSciBot are consistent, stable and under control in contrast with those returned from search engines.

Both SocSciBot and commercial search engines are useful in webometrics research. Search engines index more link pages than the SocSciBot, as they have the ability to remember old link pages and can follow link pages outside the university in question. Regardless of the differences between the two, they broadly produce similar results in academic web areas (Thelwall, 2001b; Smith & Thelwall, 2002). If the object of a study focuses on a small web area, and accuracy of the data is a high priority, or if the link counts are at different ADMs levels, then the SocSciBot is the choice. If a large web area is the object of a study, search engines can be used.

2.4 Alternative Document Models

Citations, the counterpart of hyperlinks in bibliometrics, are counted between different papers. The basic counting unit for both source and target is a scholarly paper. For instance, when counting citations between journals, if a paper in a journal is cited by two papers in another journal, the citation count of the journal is two. Even if a paper in the first journal is cited many times in different parts by a paper in the second journal, the citation count for the first journal is only one. No matter at which aggregation level the counting is taken place, such as individual, group, department, university or nation, the counting unit remains the same as a scholarly paper (Melin & Persson, 1996).

When counting links on the Web through search engines, a web page is normally the default unit. The number of links to a web page is assessed to be the number of web pages that contain at least one link to that web page (Björneborn & Ingwersen, 2004). This is different from citation counting, where the counting unit is always a paper, and different pages within one paper are regarded as one unit. The web page is not necessarily the best unit for counting links (Smith, 1999c; Thelwall, 2002b). For example, one online database may be broken down into several screen-sized pages, and links pointing to them may arise from a single motivation. The huge number of similar links may render other individually created links meaningless (Thelwall, 2004b).

The ADMs technique groups web pages from both source and target to different document levels according to their urls. The aim of the ADMs is to aggregate and remove repeated links at a higher document level. The rationale behind this is that similar web pages may be organised within one directory or domain. For example, a department's web site may have a same domain name, and a staff's web site may be within a same directory under the department domain.

Below are descriptions of the ADMs, where A is a source university and B is a target university.

- *Page ADM* At the page level, the original link data is transformed into page link data by truncating the urls of links from A to B before the first # to avoid repeated links to different parts of the same web page. Duplicate links from the same page are then removed. The link counts from A to B are the number of links in A that target B.
- *Directory ADM* At the directory model level, the original link data is transformed into directory link data, by truncating before the last slash of the urls of all source pages. Duplicated directories are merged into one directory, and repeated links within each directory are removed. Links are then counted as above.
- *Domain ADM* At the domain model level, the link data is transformed into domain link data, by truncating the source page urls after the first slash following the domain name. Then the same domains are merged, and duplicate links are removed within each domain. Links are then counted as above.
- *University ADM* For the university model, the whole university will be regarded as the unit for counting links. One university can have one link to another, if any page in A targets any page in B, otherwise none.

In addition to the above mentioned four ADMs, three 'range ADMs' have been devised by Thelwall & Wilkinson (2003b). The rationale behind these range models is: even counting links from the source site at the domain level may still produce anomalies. For example, the same person might be authorized to use multiple domains within a university and create repeated links to a target page, or people in different domains within the source site might share the information from the target site. Details of the three range ADMs are as follows.

- *Page range ADM* With the Page range model, the whole source university A is regarded as one unit, and duplicate links to the same target page are eliminated. The count of links is the same as above.
- *Directory range ADM* With the Directory range model, the whole source university A is regarded as one unit, and duplicate links to the same target directory are eliminated. The count of links is the same as above.
- *Domain range ADM* With the Domain range model, the whole source university A is regarded as one unit, and duplicate links to the same target domain are eliminated. The count of links is the same as above.

An example is given to illustrate how to count links with different ADMs between two universities: A and B. Figure 2.3 shows web pages in university A and web pages in university B. The diagram is based on Björneborn (2001a). Circles are domains, squares are web pages and triangles are directories. Each web page in A has a link to each web page in B (not shown). Table 2.2 shows the results of link counts between A and B.

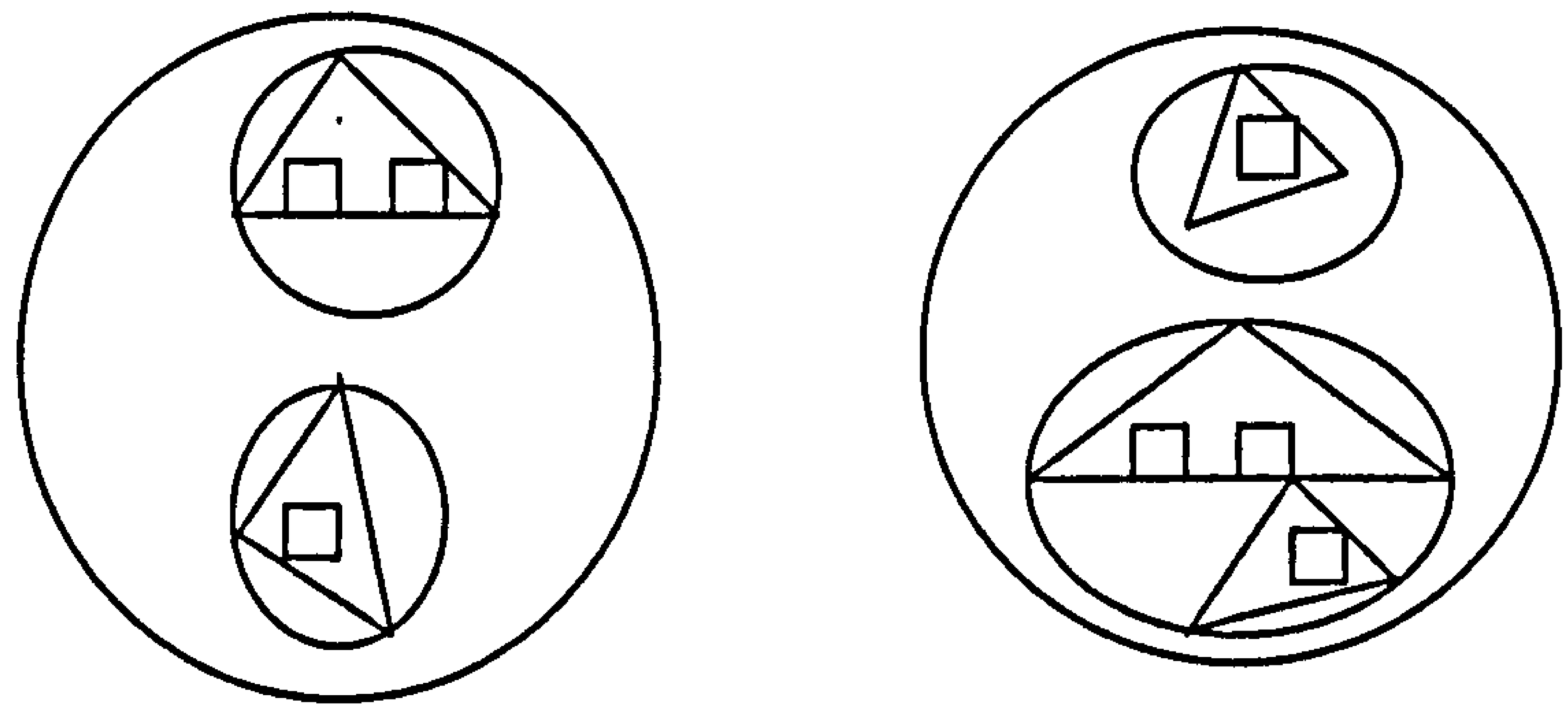


Figure 2.3 A Björneborn diagram of two imaginary web sites

Table 2.2 Number of links with different ADMs from University A to University B

Model name	Source (unit)	Target (unit)	Links from A to B
Page	Web page	Web page	12
Directory	Directory	Directory	6
Domain	Domain	Domain	4
University	Whole university	Whole university	1
Page range	Whole university	Web page	4
Directory range	Whole university	Directory	3
Domain range	Whole university	Domain	2

The advantage of the ADM technique is that large number of repeated links, which may render other links meaningless, can be effectively removed at higher aggregation document levels. The repeated links can be credit links on multiple pages of a single site, mass collections of database links or large number of links to individual resource-related web pages/sites. Links counted at directory, domain and range directory ADMs remove link anomalies more efficiently than other document models at university level (Thelwall, 2002b; Thelwall & Wilkinson, 2003b).

The ADM technique may help to trace link anomalies amongst departments. Normally when anomalies appear, no significant correlations can be identified between link counts and research measures. If the correlation between link counts and research outputs at page level is not significant, but those at directory, domain and university levels are significant, then the anomalies may well appear at the page level. Through analyzing apparent outliers in diagrams between link and research measures, reasons of link anomalies can be identified.

There are also some limitations for applying the ADM technique to count links:

1. Not every type of link anomalies can be removed. The type of anomalies caused by individual very popular web pages hosted in lower research output universities cannot be removed through any pure document model (Thelwall, 2002b).
2. Useful information may be lost while removing link anomalies. For example, two universities that are substantially linked to each other may indicate a closer relationship. At the university ADM document level, however, the link count can only be either 0 (with none links between universities) or 1 regardless of the actual number of links between them.

3. Search engines cannot be used to count links at different ADMs document levels, as they normally count links with web page as their default document unit. It is impossible to count links at different ADMs document levels in a large-scale study.
4. ADMs may not be as efficient in removing anomalies at departmental level as at university level. It is natural to organize web pages into different document units according to their urls. However, the heuristics are not as solid as per paper per document in bibliometrics, as there is no convention for the organization of urls on the Web. Within one university, some departments' web sites may be organized under individual domain names, others under different directories. It is also possible for the personal web sites from academic staff members in a department to be organized under the university domain name, even if that department has its own domain name. This will not prove to be a big problem at university level, as universities normally have many different departments, and this may average the differences out. At departmental level, this may be a big issue as a department may either be organized under a domain or a directory.

2.5 The Development and Application of the Web Impact Factor

The Web Impact Factor (WIF) is an adaptation of the Journal Impact Factor (JIF) on the Web. This is based on the analogy between hyperlinks and citations, as explained in section 2.2.

2.5.1 The Origin of the WIF

A journal that publishes more papers may receive more citations. In order to get a fairer comparison between journals, it is better to divide the number of citations a journal received by the number of papers that journal has published (Garfield, 1955). The JIF is the number of citations received in the current year by papers published in a journal during the previous two years, divided by the number of papers published within that period (ISI, 2001a). The JIF can be calculated based on the articles published in the previous one year if the field is a rapidly changing one, or based on the articles of more than two years if the field in question is less current (Garfield, 1999).

The JIF not only can help librarians to decide which journals to purchase, but also can help authors to decide suitable ones to submit to. Normally the most prestigious journals have low acceptance rates, but high JIF values. The JIF has also been used to estimate citations for a paper by granting or policy agencies, to bypass the work of counting the real number of citations (Garfield, 1999). The worth of a paper can be estimated through the JIF value, especially if the paper is new and has not yet had an opportunity to receive citations. However, this application has been criticized (Moed, 2002; van Raan, 2005).

The WIF is the application of the JIF to the Web (Rodríguez i Gairín, 1997; Ingwersen, 1998). Generally speaking, the WIF is the number of links to a site divided by the number of web pages inside the site in question. Ingwersen (1998) defined three types of WIF: internal, external and overall. For the internal WIF of a web site, the numerator is the number of internal inlinks (site selflinks); for the external WIF, the numerator is the number of inlinks counted from outside; for the overall WIF, the numerator is the number of inlinks from both within and outside. The denominators all remain the same: the number of web pages within the web site in question.

Early WIF calculations gave disappointing results (Smith, 1999c; Thelwall, 2000; Thomas & Willett, 2000; Björneborn & Ingwersen, 2001). The next section describes the problems involved in early WIF calculations.

2.5.2 Problems with Early WIF Calculations

Problems with early WIF calculations are:

1. Internal inlinks in numerator. Internal inlinks may serve as navigational tools, as mentioned in section 2.2.2. In general, the larger a web site, the more internal inlinks it will have. In contrast, inlinks from outside indicate a more definite intention to point to target pages, and therefore contain more valuable information. It is not, however, always easy to separate internal from external inlinks. For example, the School of Computing and Information Technology (www.scit.wlv.ac.uk) is a subsite of the University of Wolverhampton (www.wlv.ac.uk). Should the links from the parent site to the subsite be regarded as internal or external? As the inlinks are still within the same university web site, they are normally regarded as internal.
2. Unreliability of search engines. As discussed in section 2.3, the search engines used to collect the link data have inherent deficiencies. The results from AltaVista were especially unstable before it was re-launched in October 1999 (Rousseau, 1999; Björneborn & Ingwersen, 2001). Logically identical Boolean queries gave different results, and information scientists had to design methods to mitigate this effect (Ingwersen, 1998; Smith, 1999c).
3. Number of web pages as denominator. The original WIF denominators were the number of web pages in the target web site/sites. This includes another source of uncertainty, as there is no convention for web page output format. One document can be displayed in one web page, or separated into several screen-sized pages. For example, one online document attracts 100 inlinks from outside. If the document is represented by one huge web page, the WIF for this document is 100, while if the document is represented by 100 smaller pages, the WIF for this document will be only 1. This shows how the results can be affected by the way in which documents are presented on the Web.

2.5.3 WIF Enhancements

A number of mechanisms have been introduced to solve the problems mentioned above. Currently internal inlinks are ignored in almost all link studies. Denominators are often numbers of academic staff members rather than numbers of web pages. This removes the uncertainty about the number of web pages (Thelwall, 2001b; Smith & Thelwall, 2002). The number of academic staff member is a better representative of a university's research capacity than other measures, such as the overall budget of a university or the number of equivalent full time students etc. The WIF thus becames 'a hybrid calculation combining web information with another source' (Thelwall, 2001b).

In order to get rid of the instability of search engines, as described in section 2.3, Thelwall created SocSciBot to crawl a subset of the Web extensively and accurately. Given these enhancements, significant correlations between WIFs and non-Web research measures were found not only in the UK (Thelwall, 2001b; Thelwall, 2002a) but also in Australia (Smith & Thelwall, 2002), Taiwan (Thelwall & Tang, 2003), and Mainland China (Tang & Thelwall, 2002). Significant correlation coefficients were found mainly at the university level, but also at the departmental level (Li et al., 2003b; Tang & Thelwall, 2003b).

The significant correlations found between WIF values and research measures in different countries and at different academic levels, gave some confidence that WIF as link metric is valid to give some useful information. The validation of link metrics through correlation tests will be discussed in the next section.

Another related link metric, Web Use Factor (WUF) (Thelwall, 2003b), is the normalised number of external outlinks from a web site (Thelwall, 2003b). The only difference is that the numerator for WUF is the number of external outlinks rather than inlinks.

2.6 Correlation Tests for Link Data

A correlation test is useful to help find out whether a new data set is related to another known source (Thelwall, 2005, to appear). In bibliometrics, many correlation tests have been conducted not only between publication counts and research quality (Moed et al., 1985; Rinia et al., 1998), but also between citation counts and research quality (Cole & Cole, 1967; Cole & Cole, 1973; Oppenheim, 1995; Oppenheim, 1997; Norris & Oppenheim, 2003; Aksnes & Taxt, 2004). Oppenheim (2000) suggested several correlation tests between patent citations and other factors (e.g. renewals, sales and profits) to validate the patent citations. Similar correlation tests have been conducted between link and research measures in webometric studies with significant results (Thelwall, 2001b; Smith & Thelwall, 2002; Thelwall, 2002a). Research measures, which have been used to compare with link counts will be discussed in the next section.

Although correlation tests are useful to support the interpretations of a data set, they do not imply that one is the cause of the other. A significant correlation between two phenomena may be caused by an unrelated factor that influences both sides (Vaughan, 2001). Chen et al. (1998) found significant correlations between the thirteen Scottish computer science departments' inlinks and their organisational profiles, but the size of departments were not taken into account. Normally larger institutions have more academic staff members, receive more research funding, produce more research outputs, have better reputation and attract more external inlinks. Hence significant correlations between link counts and research outputs may be caused by the underlying size effect (Smith, 1999c).

Once the size effect has been removed through dividing both sides by the number of academic staff members, there are still some significant correlations between WIFs and research averages (Thelwall, 2001b; Smith & Thelwall, 2002; Thelwall, 2002a; Li et al., 2003b; Tang & Thelwall, 2003c). The correlation coefficient values between WIFs and research averages are even higher when only research related links are included. This may indicate that research performance is one of the reasons for universities to attract external inlinks (Thelwall, 2001b). Universities normally do not publish their research output on their public accessible web sites, and all sorts of contents can appear on the academic web sites (Wilkinson et al., 2003). Links hence cannot be used to measure research, even if only research related links are included.

The correlations between WIFs with links unrelated to research and research averages are also significant (Thelwall, 2001b). The explanation may be that the abilities of university web sites to attract inlinks are influenced directly or indirectly through the availability of computing resources or the development of technological know-how. For example, some universities with extraordinary larger WIF values relative to their research averages were found to have larger proportions of computer related academic staff members (Thelwall, 2001b). It is necessary to conduct studies on link creation motivations in order to fully understand the causation of the significant correlation results. Section 2.8 discusses this issue in detail.

Based on the significant correlation results, an association between links and research can be established. This makes it possible to disclose a perspective of research relationship amongst universities on the Web through link analysis. This is different from those identified through citation analysis which are based on formal scholarly communication. Through the motivation analysis of university interlinking, the kind of relationship is found to be informal scholarly communication (Wilkinson et al., 2003). This is not to replace but to complement those patterns identified through formal scholarly communications.

In addition to validating and interpreting link data, correlation tests have also been conducted to identify better mechanisms for link analysis. For example, the most significant correlation between link counts and research measures can be used to indicate the best ADM document model (Thelwall, 2002b; Thelwall & Wilkinson, 2003b). For example, if the correlation between domain ADM link counts and research measures is the most significant amongst other ADM document models, then this is some evidence that the domain ADM can be regarded as the most appropriate model to count links.

2.7 Evaluating Research Output

This section discusses different existing research measures, which are used to evaluate research output. In particular, various research quantitative indicators, which have been compared with link counts in previous webometric studies, are discussed and compared.

2.7.1 Existing Research Measures

Broadly speaking, there are two types of research measures, peer review and quantitative indicators. Peer review is good at judging research quality, which has a multidimensional character (Moed et al., 1985), while quantitative indicators are good to illustrate research impact, which is regarded as one aspect of research quality (van Raan, 1996).

2.7.1.1 Evaluating Research Output through Peer Review

Peer review is to judge the quality of research from individuals, research groups or programs by colleague-scientists, peers (van Raan, 2000). It is the principle mechanism to assess research quality. However, it has limitations. The major issue is its subjectivity (Moxham & Anderson, 1992). Peer judgement may be influenced by,

- Lack of awareness of quality;
- Conflict interests between reviewers and those to be evaluated (Bence & Oppenheim, 2004a);
- A bias against younger people or newcomer to a field.

In comparison, quantitative indicators, which are discussed in the next subsection, are objective and help to enhance peer review through providing reviewers additional information.

2.7.1.2 Evaluating Research Output through Quantitative Indicators
Quantitative indicators include bibliometric and non-bibliometric indicators.

1. Bibliometric indicators include,

- Publication counts
- Citation counts
- Patent counts
- Patent citation counts

The fact that an article has to go through peer review in order to be published, may suggest publication counts as quality indicators (Martin, 1996; Toutkoushian et al., 2003). However, publication counts are usually viewed as quantity rather than quality (Moravcsik, 1973). In contrast, citation counts attract more attention, since they can indicate the impact of a paper in the science community when they are applied to sufficiently large aggregates (Smith, 1981; van Raan, 1998).

Patent indicators on a macro-level measure the economic or innovative strength of a country in a certain area, but not for lower levels of aggregation (Narin et al., 2004). In comparison with citations, patent citations measure the commercial value of research rather than its academic quality. Ideally, both should be used because both are relevant. However, in reality, publication and citation counts are used more often than patent and patent citation counts for research evaluation purposes. The connection between research process and patents is not as immediate as between the research process and publications, since patent applications aim at commercial use. In addition, patent citations are more complicated than citations in publications. Firstly, patent citations are created by both the inventor and patent examiner or patent attorneys of the organization where the patent is granted (Vinkler, 1994; Meyer, 2000). Secondly, patent citations can be from either other patents or scientific publications (Ellis et al., 1978). The latter is mainly used to track the transference of technology from science rather than measure research impact (Oppenheim, 2000). Section 2.9.2.1 discusses this issue in detail.

In this context, citation counts are normally regarded as the most useful bibliometric indicators. However, there are some concerns about their application as research performance indicators. Citations do not reflect all influences on a scientific work; the research performance cannot be identified through citations in situations below:

- 'obliteration phenomenon' (Garfield, 1975). Techniques and theories become so well known that it is not necessary to cite the original sources (Oppenheim & Renn, 1978);
- Influences that appear in acknowledgements (Cronin et al., 1993).
- Informal communication (Edge, 1979).
- 'Mendel syndrome' (Garfield, 1979). Paper that is too advanced to get any citations.
- Motivations of citation creation other than endorsing the quality of a paper (Borgman & Furner, 2002). Section 2.8.1.1 discusses citation motivations in detail.
- The incompleteness of ISI databases and errors involved in it (REPP, 2003). Section 2.7.2.1 discusses this in detail.

2. Non-bibliometric indicators include,
- External funding
- Research student data
- Indicators of esteem

External funding is a well accepted research performance indicator among scientists, since the award of a grant is normally based on peer review (Tognolini & Hattie, 1994). The dispute, however, is that it is research input rather than output. Section 2.7.6 discusses this issue in detail.

Research student data has been used in various research evaluation schemes, such as the Australian funding scheme, the UK Research Assessment Exercise (RAE), and Dutch university research evaluations (Geuna & Martin, 2003). However, this measure may reflect more about teaching rather than research quality (Phillimore, 1989).

Indicators of esteem include prizes, awards, membership of professional societies, editorial board membership, membership of grant review panels, invited talks at international conferences and guest visits (REPP, 2003). However, these indicate past rather than current research performances (Phillimore, 1989), and penalise younger scientists (Wood, 1990).

2.7.2 Qualitative and Quantitative Research Measures Used in Webometric Studies

This section discusses and compares various research measures that have been used in the correlation tests in previous webometric studies. The first subsection is about bibliometric indicators, in particular, citation counts. The second focuses on peer review evaluations, especially the UK RAE scores. The third discusses the research grants as research quantitative indicators. This section illustrates the context why various research measures have been applied to this study.

2.7.2.1 Bibliometric Indicators

1. Citation counts are the most relevant data set to compare with inlinks

Publication counts, which were counted from ISI databases, were used to represent research output to compare with link counts for Australasian universities (Smith, 1999c). However, the correlation between external WIFs and publications per academic staff member was not significant. This is not surprising, as discussed in section 2.7.1.2, because publication counts are good at indicating quantity rather than quality. Significant correlations have been found between WIFs and citation counts per academic staff member in a later study (Smith & Thelwall, 2001). The citations per academic staff member were from Asiaweek (2000). Tang & Thelwall (2003c) stressed that citation counts were more relevant to compare with 'sitations' (inlinks). In that study, an 'estimated citation impact', was compared with inlinks for US chemistry, psychology and history departments. Apart from that fewer links were found amongst the history departments, significant correlations were found between inlinks and 'estimated citation impact' for both chemistry and psychology departments. A department's 'estimated citation impact' was multiplying ISI citation impact factor in relevant field of a university by the publication counts for that department. The ISI citation impact factor (1997-2001) was from ISI Contract Research Services, while the publication counts were from ISI databases in 2000.

As discussed in section 2.7.1.2, citation counts are normally the best single bibliometric indicator for measuring research performance. Recall, that it is the structural similarity between citations and hyperlinks that triggered the webometric studies. As a result, the most relevant research quantitative indicators to compare with link counts are citation counts. For example, the following related pairs could be used in the correlation tests:

- Links versus citations;
- WIFs with web page counts as denominators versus citations per publication (CPP);
- WIFs with staff member counts as denominators versus citations per staff member

2. Counting citations from ISI databases

Eugene Garfield is the very person in the history of citation analysis. He published the seminal paper on citation indexes in 1955 (Garfield, 1955). He not only promoted the citation concept in research but also used it to create a business by establishing the Institute for Scientific Information (ISI). The ISI (2005) has now indexed approximately 8,700 of the most prestigious, high impact research journals in the world. It includes three databases: Science Citation Index (SCI), Social Sciences Citation Index (SSCI) and Arts & Humanities Citation Index (A&HCI). These databases make it possible to count citations automatically over a large scale for different disciplines.

Generally there are two major approaches for counting citations at departmental level. One is querying academic staff members' names through the Web of Science (Oppenheim, 1995; Holmes & Oppenheim, 2001; ISI, 2001b; Norris & Oppenheim, 2003), and getting the citation counts for each department by accumulating citations that its academic staff members receive. Another method is a semi-automatic approach to get the relevant citation counts over ISI's databases by matching terms in the address part of each publication record (Moed et al., 1995; Bourke & Butler, 1998; Toutkoushian et al., 2003).

Regardless of the errors contained in the ISI databases (Moed, 2002; van Raan, 2005), the first approach can count all citations indexed by the ISI for the academic staff members within a department, even if the publications are in different disciplines (e.g. staff members in physics department may publish papers in mathematics journals). The limitations for the approach are: firstly it may involve citation counts from authors with same surname and initials in other departments or even other countries; secondly it is only suitable for a small-scale citation count, since to collect the citation counts manually for a large number of staff members can be very error prone and time consuming.

The second approach is more plausible to count citations in a large scale. The country name of the address part is well unified by the ISI. However, the situation is different under the country level. One university or department may have a variety of names in the ISI databases (Moed et al., 1995; van Raan, 2005). Thus the drawback is that a great deal of effort is needed to clean and unify the addresses of universities or departments in order to collect accurate citation counts for them. In order to avoid this, a compromised approach can be to count citations by matching part of an address which is unique to each department. For example, postcodes can be used for this purpose. The possible drawback for this is that not every address contains a postcode. As a result of this, some citations are simply lost. This should not be regarded as a problem if the data set is big enough.

Issues for counting citations from ISI databases are as follows:

- Coverage of publications. With regard to the coverage of publications, ISI claims to index the core journals in all research fields. As a result, its coverage of natural science is better than that of social science and humanities, as the latter relies heavily on publications other than journal papers, e.g. books, monographs and book chapters which are not significantly covered by ISI (Glänzel & Schoepflin, 1999; van Raan & van Leeuwen, 2002; Moed, 2005). This can be a big issue if citations are counted for departments in social science and humainites.
- Mistakes within the ISI databases. Mistakes have been identified in ISI databases, as they are created for information retrieval rather than counting citations accurately (Wouters, 1999). The average rate of non-matching citing-cited references is about 7% (Moed, 2002). Errors in authors' names especially occur more often in non-English speaking countries (van Raan, 2005). When counting citations for research evaluation purposes, these errors should be corrected (van Raan, 2005).
- Publication types. With regard to publication types, papers, letters, notes and reviews have been indexed in ISI databases.
- Subject classification. The ISI classifies journals into different categories. Papers are classified according to journals that they belong to. A journal may cover more than one discipline and be allocated into different categories and may lead to repeat citation counting (Glänzel, 2003). Repeat journals for each discipline have to be removed before the citations are counted.
- Multiple authorships. Despite the fact that one single authored paper in quantity equals to one multi-authored paper, it is difficult for counting citations to multi-authored papers, since a decision has to be made

whether the citations should be counted equally to each author or differently (Lindsey, 1980; Long & McGinnis, 1982). There are three approaches for this (Cronin & Overfelt, 1994):

- 'straight counts', citations are credited to only first authors.
- 'whole counts', citations are credited equally to each co-author the paper receives.
- 'adjusted counts', citations are allocated to each co-author according to their contributions to the paper.

The 'adjusted counts' approach is appealing. However, it is difficult to identify a consistent rule for it (Cronin & Overfelt, 1994; Laudel, 2002). Although 'straight counts' illustrates the importance of first authors' contributions, it misses information of those junior staff members who normally are not first authors. In one study, the citations from 'straight counts' and 'whole counts' correlated significantly (Lange, 2001).

- Self-citations. Self-citations are often included in the citation counts, since they usually indicate that authors build on their previous work (Katz & Hicks, 1997), and also Glänzel & Thus (2004) find out that at macro level, there is no need to exclude self-citations.
- Language preferences. ISI covers more English journals than non-English (van Raan, 2005). Citation counts may be lost to those non-English journals with various names (Ren & Zu, 2002).

2.7.2.2 Ranking Schemes

1. Different ranking schemes applied in previous correlation tests in webometrics.

- AsiaWeek (2000): The Australasian universities' rankings from AsiaWeek were compared with university WIFs with no significant result (Smith & Thelwall, 2001). The reason may well be that the ranking scheme is not for all the Australasian universities, but only the top ones. The ranking scheme was based on data from 1999 questionnaires and supplemented by other sources. The overall score for each university was calculated by values including 'academic reputation' (20%), 'student selectivity' (25%), 'faculty resources' (25%), 'research' (20%), 'financial resources' (10%). Weight of each value is in brackets.
- NetBig (2001): The ranks of Chinese universities from NetBig's 'Comprehensive Ranking 2001' were compared with relevant number of inlinks with significant correlation coefficient values (Tang & Thelwall, 2002). The NetBig ranking scheme incorporates multiple values including: 'academic reputation' (15%), 'academic status' (20%), 'academic achievements' (22%), 'student performance' (12%), 'faculty resource' (19%) and 'facility resource' (12%). Weight for each value is in brackets.
- RAE (HERO, 2001d): The UK Research Assessment Exercises (RAE) scores have been used to compare with inlink metrics in many previous studies with significant results (Thelwall, 2001b; Thelwall, 2002a; Li et al., 2003b). The RAE operates through a process of peer review by experts of high standing covering all subjects. The experts will evaluate research excellence based on the information submitted: 1) 'staff information', with regard to research-active and non research-active, 2) 'research output', up to four items of research output for each researcher, 3) 'textual description', with regard to

information about research environment, strategies and qualitative information on research performance, 4) 'related data', including research funding, research students, research studentships, research degrees awarded and indicators of peer esteem.

The results from both AsiaWeek and NetBig are comprehensive university rankings with regard to various aspects, whereas the RAE focuses on mainly research excellence. In comparison, the UK RAE scheme is the most authoritative one, since it determines the allocation of billions of government research funding (HERO, 2001c), while AsiaWeek and NetBig ranking do not connect to any funding allocation, but only supply information for their audience. Besides, both AsiaWeek and NetBig are organized at university level, while the RAE is organized at departmental level. As a result, only the RAE results can be used to represent departmental research performances. The next part of this section will discuss the UK RAE ranking scheme, especially the issues involved when converting the RAE scores into research quantities through a formula.

The advantage of using ranking scheme to represent research performance is that it is normally publicly available information, and no extra efforts are needed in order to collect them. The drawback, however, is that they are not research quantitative indicators, the comparison between rankings and inlinks are not as relevant as citation counts.

2. The calculation of RAE averages

The Research Assessment Exercise (RAE) is an extensive exercise to evaluate the quality of research conducted in the UK universities by the Higher Education Funding Councils. Since its inauguration in 1986, there have been altogether 5 events, taking place consecutively in 1989, 1992, 1996 and 2001 (Bessant et al., 2003; Geuna & Martin, 2003; Bence & Oppenheim, 2004b; HEFCE, 2005). The main purpose of the RAE is to direct the UK government research block grant to universities according to their research qualities. The RAE 2001 is the most recent, and the next RAE will be in 2008.

RAE 2001 includes 69 Units of Assessment (UoA) (HERO, 2001b), which cover different disciplines. The results were reported in seven ratings in each UoA. The ratings of research quality were defined as follows (HERO, 2001a):

1: Quality that equates to attainable levels of national excellence in none, or virtually none, of the research activity submitted.

2: Quality that equates to attainable levels of national excellence in up to half of the research activity submitted.

3b: Quality that equates to attainable levels of national excellence in more than half of the research activity submitted.

3a: Quality that equates to attainable levels of national excellence in over two thirds of the research activity submitted, possibly showing evidence of international excellence.

4: Quality that equates to attainable levels of national excellence in virtually all of the research activity submitted, showing evidence of international excellence.

5: Quality that equates to attainable levels of international excellence in up to half of the research activity submitted and to attainable levels of national excellence in virtually all of the remainder.

5*: Quality that equates to attainable levels of international excellence in more than half of the research activity submitted and attainable levels of national excellence in the remainder.

Since a department may submit in more than one UoA and has more than one rating score, an average RAE score is necessary to represent the department's research performance. A technique was developed by the Times Higher Education Supplement (Mayfield University Consultants, 2000) to compute university average RAE scores across all Units of Assessment (UoA) from a university. Influential league tables in newspapers to guide students in their choice of universities are created through the same calculation (EducationGuardian, 2001; THES, 2005). This technique has been applied to calculate average RAE scores not only for universities but also for departments in previous webometric studies (Thelwall, 2002f; Thelwall, 2002b; Li et al., 2003b). The formula below shows how it works.

$$RAE_i = \frac{\sum_j staff_{i,j} \times grade_{i,j}}{unsubmitted_i + \sum_j staff_{i,j}}$$

Here $staff_{i,j}$ is the number of full-time equivalent staff submitted by university i to unit of assessment j, $grade_{i,j}$ is the grade allocated to university i for unit of assessment j, changed to a seven point scale (from 1, 2, 3b, 3a, 4, 5, 5* to 1, 2, 3, 4, 5, 6, 7), and $unsubmitted_i$ is the full-time equivalent staff at university i that were not submitted to any UoA. Note that the denominator is the total eligible full-time equivalent researching and teaching staff in university i, including all staff not submitted to the RAE. It is possible to estimate a university's total research productivity as the numerator part of the formula. By doing this, the research qualities, in RAE terms, are converted into research quantities.

The calculation needs some justification because it is not self-evidently meaningful. According to 'A Guide to the 2001 Research Assessment Exercise' (HERO, 2001c), grade 1 and 2 normally receive no funding, while grade 5* attracts approximately four times as much funding as grade 3b for the same number of research active staff members submitted. As discussed in section 2.7.1, research funding can be used as research performance indicator. One can see that the RAE grades are not linear, and the differences between grades are unknown. This raises a critical issue as whether the RAE scores can be mathematically manipulated in this way. As discussed in section 2.6, significant correlations between inlink counts and research measures indicate the association between inlink counts and research performance. Significant correlations both between research productivity and inlink counts and between RAE averages and

WIFs in previous webometric studies (Thelwall, 2002f; Thelwall, 2002b; Li et al., 2003b) may serve as some evidence that the formula above is reasonable to reserve the relative research quantities of the universities and departments.

Recall, that research income can be used as research performance indicator. The research income each department received was included in the RAE 2001 submission (HERO, 2001e). It is possible to conduct the correlation tests between the research productivity (the numerator part of the formula) and research income, and between RAE averages and research income averages, in order to further justify the formula. If the results were positive, the formula should be acceptable as proxies of research quantitative indicators in the correlation tests in webometrics.

2.7.2.3 Research Grants

Research grants used in previous webometric studies to represent research performances:

- Research Quantum (RQ) (DETYA, 2001): Significant correlations were found between RQ per member of staff and various WIFs for Australian universities (Smith & Thelwall, 2002). RQ is an official Australian government research infrastructure grants, which is similar as the UK's RAE in ultimate objective, but is calculated through a simple publicly available formula rather than peer review. In that formula, previous research funding received weighted 80%, while research output like number of publications and postgraduate degrees completed weighted 20%. The RQ has been replaced by Institutional Grants Scheme (IGS) in 2002 (DEST, 2005). The IGS funding is calculated by a so-called performance-based formula, whereas research income weighted 60%, publications weighted 10% and higher degree research students weighted 30%. The funding scheme was criticised as encouraging research quantity rather than quality (Butler, 2003).

- NSERC (2003): Vaughan & Thelwall (2005) employed NSERC research grants as a variable to indicate faculty research measures, in order to predict the number of inlinks for Canadian universities. The NSERC research grants is a national funding scheme organised by National Sciences and Engineering Research Council of Canada for Canadian higher education. It is based on peer review over applications of various projects. The results were organized not only by universities but also by disciplines. It is possible to obtain the research funding for each department in Canada. Thus the funding scheme is a good candidate to validate departmental link data in Canada.

Research funding obtained from external sources is one of the preferred quantitative indicators of universities' research performances (Tognolini & Hattie, 1994; Martin, 1996). The reason is that the award of a grant is normally based on peer review (Gillett, 1991; Hornbostel, 2001). On the other hand external funding is a complement indicator for governmental research evaluations (HERO, 2001c; DEST, 2005; TEC, 2005).

Nevertheless, there are some objections to research grants as research indicators:

- Research grants measure research potential rather than achievement (Gillett, 1991)
- Research grants are research input rather than output indicators (Phillimore, 1989; Johnes & Johnes, 1995)

2.7.3 Summary

Research quality is traditionally measured through peer review, while research impact is measured through quantitative indicators. In comparison,

- Peer review is only suitable for a small scale, while quantitative indicators are suitable for large scales;
- Peer review is subjective, while quantitative indicators are objective;
- Quantitative indicators are good complementary information for peer review.

No known quantitative indicators can replace peer review to measure research quality. However, various quantitative indicators have attracted more and more attentions, especially bibliometric indicators such as citation counts from ISI's databases. At least when quantitative indicators gave contradict results than peer view, it is worth to have another look to make sure the results are accurate.

Various research measures have been employed in webometric studies, particularly in correlation tests where they are compared with link counts. Citation counts are regarded the most relevant research measures for this purpose, because that

- Citation counts are the most used bibliometric indicators;
- It is the analogy between citation and link that triggered the webometric studies;
- When other research measures are not available, citation counts can always be collected from ISI's databases, although with some problems.

Other research measures such as peer review and research grants have also been employed for the same purposes. In comparison, they are not as relevant as citation counts; and if they are organized by universities, they cannot be used in correlation tests at departmental level. Nevertheless, it is cheaper and more convenient to collect those public accessible peer review results or research grants than counting citations from ISI's databases. One can be more confident about the relationship between research performance and link counts, if significant correlations have been discovered between links and various research measures from different aspects.

2.8 Link Motivation Analysis

As discussed in section 2.6, significant correlations between links and research measures do not imply causation. The motivations for link creation should be analysed to reveal the nature of links in order to better understand the underlying relationship. In bibliometrics, the nature of citations has been investigated extensively for decades, although without a definite conclusion. In this section, the first subsection reviews various approaches for citation motivation analysis; the second subsection discusses those for link motivation analysis.

2.8.1 The Study of the Nature of Citations

Citation studies have been used for various purposes, such as evaluating research outputs, as discussed in section 2.7; investigating the history of science (Baird & Oppenheim, 1994); mapping science structure (Small, 1998); identifying patterns of communication amongst authors, journals, institutions or nations (Borgman & Furner, 2002). Cronin (1984) pointed out that one should not only focus on citation analysis but also the reasons, 'Why does author a decide to link document x to document y?'. If one can understand the nature of citations, one can better understand what citation analysis signifies. The first part will review reasons for citations; the second part will discuss various techniques applied to identify citation motivations.

2.8.1.1 Reasons for Citations

Generally speaking, the citation studies are based on two assumptions that citations are created for two types of reasons.

- 'Normative': used to acknowledge intellectual debts to cited documents. Normally the cited documents are relevant to the citing document's topic, and provide useful background (Cronin, 1984; Liu, 1993; Case & Higgins, 2000; Borgman & Furner, 2002).
- 'Persuasion': used as a tool to support citing authors' claims (Gilbert, 1977).

Garfield (1996) gave guidelines on 'when to cite', based on three years' investigation of graduate student citing behaviour. According to him, citation should be used as a means for the 'acknowledging of intellectual debts'. Credit should be given to earlier contributions to authors' work. If all authors cited the citation worthy works (papers of high quality reviewed by peers), and all citation worthy works are cited by authors, then the quality of a given citable work can be evaluated by the number of citations received by it. In fact, this is the presumptive condition of citation analysis, the 'normative' style (Merton, 1973). This is, however, criticized by the argument that credit-worthy papers might have high levels of creativity or innovation, yet have no relevant relationship with any potential citing papers (Shadish et al., 1995; Borgman & Furner, 2002).

Garfield (1965) listed fifteen reasons for citation, which are based on observation and anecdote, to warn against 'indiscriminate' or 'unqualified' use of citation counts. These are:

- Paying homage to pioneers
- Giving credit for related works
- Identifying methodology, equipment, etc.
- Providing background reading
- Correcting one's own work
- Correcting the work of others
- Criticising previous work
- Substantiating claims
- Alerting to forthcoming work
- Providing leads to poorly disseminated, poorly indexed, or uncited work
- Authenticating data and classes of fact - physical constants, etc.
- Identifying original publications in which an idea or concept was discussed

- Identifying original publications or other work describing an eponymous concept or term
- Disclaiming work or ideas of others (negative claims)
- Disputing priority claims of others (negative homage)

In reality, citers' motivations and goals can be more complicated than the 'normative' or 'persuasive' assumptions, and can be influenced by citer's perceptions, attitudes, prejudices or erudition (Cronin, 1984). Authors are found to make citations carelessly and to be biased in their citing habits (MacRoberts & MacRoberts, 1989). When analysing the nature of citations, various factors should be taken into consideration (Baird & Oppenheim, 1994). Empirical means have been applied to identify the nature of citations, either through content/context analysis or citer motivation surveys (Liu, 1993; Case & Higgins, 2000; Borgman & Furner, 2002). These studies focus on either the attributes of cited papers, or that of pairs of citing and cited papers, which are the reasons for authors to create citations.

Borgman & Furner's (2002) review chapter summarised the attributes of both cited papers and cited/citing pairs that influence citations as follows:

- Attributes of cited papers attracting more citations:
  - i. Quality of content, higher quality papers (Shadish et al., 1995).
  - ii. Sex of author, papers written by male authors (Baldi, 1998).
  - iii. Number of authors, papers written by a large number of co-authors (Rousseau, 1992).
  - iv. Source, papers are published in journals (Baldi, 1998).
  - v. Citedness, papers have been cited many times (Merton, 1968a).
  - vi. Subject, papers are recent and on a hot topic (Seglen, 1998).
  - vii. Approach, review papers.
  - viii. Field, papers in basic research rather than applied
  - ix. Assimilation: papers do not cover material that is now so well-understood that it has been 'obliterated by incorporation' (Merton, 1968b).
- Attributes between citing and cited pairs that attract citations:
  - i. Relatedness of content, the content of cited and citing papers is related (White & Wang, 1997).
  - ii. Field, the field in which the pair of papers is published has high citation rates (Vinkler, 1996).
  - iii. Persuasiveness, the cited paper is viewed by the citing author as (Case & Higgins, 2000):
    - Supportive to the citing paper.
    - Bringing authoritative to the citing paper due to the fame of the cited author.
    - Meeting the expectations of the citing document's audience.
  - iv. Availability, the cited paper is available to the citing paper. For example, free online availability substantially increase a paper's citations (Lawrence, 2001).
  - v. Author self-citation, between a pair of cited and citing papers, at least one cited author is in the list of citing authors (Snyder & Bonzi, 1998).

vi. Journal self-citation, the pair of papers are published in the same journal.
vii. Social citation, the citing author is related with the cited one, such as they are friends, colleagues, co-authors, etc., or the cited author is an editor or referee for the journal the citing author intends to submit the paper (Liu, 1993).
viii. Language self-citation, the cited and citing pairs share the same language (Yitzhaki, 1998).
ix. Nationality self-citation, the cited and citing authors are from the same country (Herman, 1991).
x. Time difference, the publication time difference between the pair of cited and citing papers is not long.

These lists show the influence of factors other than the direct relevance of the cited document, and hint at the underlying influence of informal scholarly communication.

2.8.1.2 Techniques to Identify Citation Motivations

There are two main approaches in this area. One is through citation content/context analysis (Small, 1983; Liu, 1993), in which a classification scheme is devised, to disclose the reason why a citation exists. The other is through interviewing, or questionnaire survey with authors directly about their reasons for citing (Case & Higgins, 2000).

- Content/context analysis

The first in-depth study of the quality of citations and the context in which citations are made, was conducted by Moravcsik & Murugesan (1975). The authors indicated two main reasons for lack of in-depth citation studies at that time. The first was that the use of citations was 'relatively recent'; while the second was that researchers who had conducted similar work did not 'understand the technical scientific content of the papers they handle', and thus could not catch the subtleties of citations which are 'connected with the quality of the paper cited, and the context in which the citation is made' (Moravcsik & Murugesan, 1975). They devised a classification scheme to identify the nature of a citation. In that study, 30 papers dealing with theoretical high energy physics, are selected from Physical Review (1968-1972). Four questions were asked to help classifying references found in these papers. Is the reference

1. Conceptual or operational?
2. Organic or perfunctory?
3. Evolutionary or juxtapositional?
4. Confirmative of negational?

The study found that two-fifths of the references were 'perfunctory', and one-seventh were 'negational', which raised serious doubts about 'evaluative citation analysis'. Cano (1989) adapted Moravcsik & Murugesan's model and devised eight types of citations by their locations within 42 papers. In that study, 26% of the citations were 'perfunctory', whereas only 2% were 'negational'. Over a third of the citations in the introduction sections were 'perfunctory'.

An alternative approach, which was devised by Chubin and Moitra (1975), was a six-class citation topology. It distinguishes affirmative citations from negative ones, 'essential' from 'supplementary', 'basic' from 'subsidiary'. Based on Chubin and Moitra's work, Oppenheim & Renn (1978) investigated reasons why 23 highly cited old papers in physics, which were published between 1896 and 1921, still heavily cited by then. The authors tracked papers published in 1974 and 1975, which had cited those 23 papers, and identified the reasons for them. The classification scheme is as follows:

A. Historical background
B. Description of other relevant work
C. Supplying information or data, other than for comparison
D. Supplying information or data for comparison
E. Use of theoretical equations
F. Use of methodology
G. Theory or method not applicable or not the best one

The main conclusion of the study was that 40% of the citations to the old papers was historical, while 60% of the citations showed that the old papers were still highly cited since they were still relevant in the citing papers. Less then 2% of the citations were 'partially negational' citations.

In comparison, Moravcsik & Murugesan studied a set of citing papers to classify the citations in them, while Oppenheim & Renn focused on a set of cited papers, discovered the citing papers and classified reasons why they were cited.

There is much literature on citation content/context analysis, which devised classification schemes to understand the relationships between cited and citing papers (Liu, 1993). This type of approach is based heavily on inference rather than empirical evidence, which 'involves a large degree of personal judgement as well as an in-depth knowledge of the subject matter' (Peritz, 1983). The citers' motives, which were judged through this approach, may not have been in their mind at the time of citing. Citations, being a personal practice, should be studied through consulting the citers themselves regarding reasons of citation creation (Shadish et al., 1995).

- Citer motivation survey

The first direct systematic survey on authors about their citing motivations was conducted by Brooks (1985). He reviews various models of citation behaviour and identified seven common reasons of citing as follows:

1. Currency (providing up-to-date information)
2. Negative credit (cited for criticism or correction)
3. Operational information (a concept or theory is referred)
4. Persuasiveness (cited for convincing peers)
5. Positive credit (cited for paying homage and giving credit)
6. Reader alert (alert the reader new or obscure sources)
7. Social consensus (cited for an unspecified and vague perception of consensus in a field of study)

Brooks surveyed 26 authors for their citing motivations in their recent publications. He found out that the 'persuasiveness' appeared the major motivator, 'negative' credit was very weak. He also found out that citer motivations were complicated, and a document may contain many items of information that may be cited for a number of reasons (Brooks, 1986). Bonzi & Snyder (1991) surveyed 51 self citing authors in several natural science disciplines. The study found out that there were few differences in reasons for self citing or citing to others' works.

The studies conducted by Brooks (1985), Bonzi & Snyder (1991), Shadish et al. (1995) and Case & Higgins (2000) all employed self-administered questionnaires. Another approach is to interview authors directly for their citing motivations (Chubin & Moitra, 1975; Cronin, 1984). Shadish et al. (1995) pointed out that the combination of interview and questionnaire survey may better disclose authors' citation motivations than only questionnaire survey, where the analyser can have a chance to observe authors reactions during the interview. The study conducted by White & Wang (1997) employed exactly the combined approach. They investigated 12 agricultural economists about the way they cited, and identified 27 criteria for the respondents to cite documents. White & Wang argued that intensive interviews can uncover citers' motivations which can not be identified through either surveys or content/context analysis. However, compared with other approaches, interviews involve a great deal of time and effort. Furthermore, through face-to-face interview, citers may feel under pressure and may reply to questions not as honestly as when dealing with questionnaire survey.

Even though interview or survey can better disclose citers' motivations, they are more expensive to conduct. This may be the reason why more experiments have been conducted through content/context analysis.

2.8.2 The Study of the Nature of Hyperlinks

Many link analysis studies based on citation analysis techniques have been conducted since 1996 (Larson, 1996), for example, measure web impacts (Ingwersen, 1998; Smith, 1999b; Thelwall, 2001b), mapping link structures (Thelwall, 2001a; Thelwall & Wilkinson, 2003a), identifying informal scholarly communication amongst various academic web sites (Thelwall, 2002d; Thelwall, 2002c; Thelwall & Tang, 2003). In comparison, less is known about why hyperlinks are created. The reasons are: firstly the use of hyperlinks is more recent than that of citations; secondly, online publication is anarchic, and various formats exist; thirdly, it is complicated to extract relevant information due to technical complexity. This section reviews link motivation studies in the context of citation motivation analysis, which is discussed in the last section. The first part will review reasons of hyperlinks, while the second part will discuss various approaches employed to identify link motivations.

2.8.2.1 Reasons for Hyperlinks

The appearance of e-journals blurred the boundary between traditional and web publications. Kim (2000) found that most motivations for hyperlinks in e-journals are 'grounded in conventional citation practices', and extended to link to multimedia and other resources. Students tend to cite internet sources in addition to traditional publications (Oppenheim & Smith, 2001).

As a whole, the motivations for academic web links are different from those for citations since links have additional functions other than referring to articles, e.g. links to pictures, videos or software downloading etc. Compared with citation motivations, link motivations for academic web pages

- Are more trivial (Thelwall, 2003c). Academic web sites contain unscientific and apparently unnecessary material. A lot of links between 25 UK university web sites were for informal and recreational purposes (Thelwall, 2001b). Only two out of the 414 academic links were equivalent to journal citations (Wilkinson et al., 2003). Even when the links were limited to a sample of research related web sites, only 20% could be regarded as research links analogous to citations (Smith, 2004). Based on a qualitative analysis on a sample of 100 random UK university inter-site links, which point to university home pages, Thelwall (2003c) postulated four types of link motivations: 'ownership' for links acknowledging authorship or co-authorship, 'social' for links with a primarily social reinforcement role, 'general navigational' for those with a general information navigation function and 'gratuitous' for those that serve no communication function at all. The 'gratuitous' is also regarded as 'web convention' where the author of source page is expected to insert, often without the intention that the link is to be followed (Bar-Ilan, 2005).
- Are for navigational purposes especially for site self-inlinks (Smith, 1999a; Björneborn & Ingwersen, 2001), even if Bar-Ilan (2004b) argued that the percentage of content-bearing self-inlinks is significant and is not negligible. Haas & Grams (2000) examined all links on 75 pages and found that the links were for: navigation, expansion, resource and miscellaneous.
- Include accessing multimedia resources (Kim, 2000).
- Have disciplinary differences since various disciplines use electronic media differently (Kling & McKim, 2000). Universities have more computer science staff members tend to attract more inlinks (Thelwall, 2001b). Scientists make more use of the Web than social scientists (Thelwall & Tang, 2003). Science and engineering dominate the UK university web sites (Thelwall & Price, 2003). Library information science journal web sites attract more inlinks than those of law (Vaughan & Thelwall, 2003). The US chemistry and psychology departments interlink more than history departments (Tang & Thelwall, 2004). Computing web sites are the most linked in Australia and Taiwan (Thelwall et al., 2003c). Bar-Ilan (2005) also confirmed 'the level of technology enabling and background influences web publishing'.
- Tend to access a wide range of information, rather than providing specific content (Thelwall, 2002g). UK library and information departments were found to attract more inlinks if their web sites contain a wide range of types of material (Thomas & Willett, 2000). The top 100 linked to pages on UK university web sites were dominated by university home pages (Thelwall, 2002g). Chu (2005) found that 73% links to the ALA accredited library information science schools were created for accessing service or home page. Through investigating inter links amongst the 8 Israeli academic institutions, Bar-Ilan (2004a) found that the source pages are dominated by link lists,

while target pages are dominated by homepages. Smith (2004) studied the links to a sample of research oriented web sites, and found that the most common source page was directory or subject guide, while the most common target pages was the main page (homepage) of an organization. He concluded that 'the most common reason for linking was the provision of further information to amplify the content of the source page'.

- Are rarely negative (Chu, 2005).

Subject inter- and intra-linking have also been studied. Similar subjects interlink more than those from dissimilar subjects (Thelwall et al., 2003a). Links created from maths, physics and sociology subjects are different in both inter-subject and intra-subject interlinking (Harries et al., 2004). In addition to such disciplinary differences, linguistic (Thelwall et al., 2003b) and geographic (Thelwall, 2002c) are two facts that affect link creations. Both the site age and content are significant factors for law and library information science journal web sites to attract inlinks (Vaughan & Thelwall, 2003).

Wilkinson et al. (2003) studied 414 links from ac.uk domain, and discovered over 90% were created for broadly scholarly reasons. They concluded that a range of informal types of scholarly communication dominated academic web links. Thus the link patterns identified through link analysis can be used to illustrate informal scholarly communications (including education-related).

2.8.2.2 Techniques to Identify Link Motivations

It is not always possible to identify authors for web pages. Even if webmasters can be identified, perhaps they may not be the persons who determine the texts or links on web pages. As a result of this, the interviews or questionnaire surveys of citers have been ignored in link motivation analysis. However, the content/context analysis has been applied extensively through various classification schemes.

Cronin (1998) devised 11 categories to classify genres of invocation of names on the Web for five highly cited library information science professors. Although not a link study, it is important for its methods and typology. The eleven categories were:

- Abstract
- Article
- Conference proceedings
- Current awareness
- External home page
- Listserv
- Personal/parent organization home page
- Resource guide
- Book review
- Syllabus
- Table contents

Five search engines were used to search the five names. Relevant web pages, where the names were invoked (not necessarily with links), were classified according to the categories. The invocations were found to 'lump together a disparate set of rhetorical genres'. These have the potential to disclose the total impact of a scholar's 'ideas, thinking, and general professional presence'. Although the study was about 'invocations' rather than 'hyperlinks', and it did not take the context (in which invocations occurs) into consideration, it was the first time that such an approach was conducted. Later, Wilkinson et al. (2003) based on this approach studied UK academic links to identify the motivations of link creations. A link structure of 107 UK universities, which was crawled by the SocSciBot in July 2001, was used to extract the sample. Firstly, 20 external outlinks, which target at an ac.uk site, were selected randomly from each university's link files. Then the first 550 links were used in the study. The first 50 links were used to devise the classification scheme. The scheme was a context analysis. Through visiting both source and target pages, the context were then determined. The rest 414 valid links were classified according to the scheme. The 10 categories of reasons of linking were:

1. Student learning material
2. Information for students
3. Research support and resources
4. Research partners
5. Recreational
6. Page creator or sponsor
7. Research reference
8. Tourist information
9. Libraries & e-Journals
10. Similar department

However, the classification proved to be very thorny, since the range of potential reasons available, unreliability of objectively inference link motivations and potential multiple motivations. As a result, three broader categories were created as follows:

A. Student learning material (old category 1)
B. Student/staff support (old categories 2,3,4,6,7,9,10)
C. Non-academic (old categories 5, 8)

Wilkinson et al. found out that over 90% of the academic links were created for scholarly reasons, whereas only 2 were citation equivalent. This implied that link counts can be used to indicate an 'agglomeration of connections related to scholarly activities in a wide variety of ways'. This also showed that the overlap between hyperlinks and citations was very small, even on university web servers. However, disciplinary link use differences were not considered in that study.

The classification schemes have been focused on target pages (Thelwall, 2001e); source pages (Thelwall, 2003c); both source and target pages (Harries et al., 2004); link contexts (Wilkinson et al., 2003; Bar-Ilan, 2004b; Chu, 2005); Source and target pages together with link context (Bar-Ilan, 2004a; Smith, 2004; Bar-Ilan, 2005). Unfortunately the results are not directly comparable. Ideally, a standard method should be adopted.

The analysis of target pages is good at disclosing which types of pages are of interest in a set of link data. However, Bar-Ilan (2005) argued that 'the characteristics of the source pages are more influential than that of the target page' when examining linker motivation. The analysis of both source and target pages show which type of page makes more links, and which type of page is more likely to be targeted. The distribution of page types is different between source and target pages (Bar-Ilan, 2004a; Harries et al., 2004; Smith, 2004). For example, directory or subject guide pages are more likely to be source pages, while homepages are mostly targeted (Thelwall, 2002g). Link context analysis is an analogy for citation context analysis. The most comprehensive analysis is to analyse source and target pages together with the link context. A more complete picture of link creation motivations can be revealed. However, this involves considerable effort.

Harries et al. (2004) investigated intra- and interdisciplinary links through analysing links created from maths, physics and sociology domains in the UK university link data. The classification scheme was from three different aspects: content types, genres and owners.

Content types:
- Research publication
- Research description
- Research activities
- Subject information
- Teaching
- Support
- Documentation
- Recreation
- Other/Unknown

Genre types:
- University home page
- Department home page
- Research group home page
- Academic home page
- Other home pages
- Link list
- Other
- Unknown

Page owners:
- University
- Department
- Research group
- Academic
- Student
- Outside University
- Other

- Unknown

They found not only significant difference in hyperlinks between subjects but also differences between intra- and interdisciplinary links. In that study, almost 30% of source or target pages were research descriptions or research publications, and nearly half were subject information. They also found that almost half of the source pages were forms of link list, while various types of home page were common targets.

The most comprehensive classification scheme was devised by Bar-Ilan (2004a; 2005) in two very similar studies of inter-university links amongst 8 Israeli academic institutions. Source and target pages together with links were classified in 11 facets to characterize the link structure between universities. The subcategories between the two studies were slightly different. The main categories for source pages were:

- Link context
- Page type
- Intention of page
- Creator/publisher
- Academics field of creator
- Languages

In contrast, the subcategories of 'page type' are similar to Harries et al.'s page genre types; those of 'intention of page' are similar to page content types; those of 'creator/publisher' are similar to page owners. The 'academics field of creator' is about the discipline that a page belongs to. 'Link context' is about the anchor text and the text surrounding it. This is a new category compared with other approaches, and has the potential to disclose more information about the motivations of links. As it is in the source page, the subcategories of 'link context' are the same as 'page type' (Bar-Ilan, 2005). 'Languages' were not used in Harries et al.'s study, since this category is more relevant for non-English speaking countries.

The main categories for link were:

- Relationship between source and target
- Intention of link
- Tone of link
- Placement of link

The subcategories of 'intention of link' are the same as 'intention of page', which are similar to the link categories in some previous studies (Wilkinson et al., 2003; Smith, 2004; Chu, 2005). The rest three categories are unique, which can help to better understand the nature of links in various perspectives. 'Tone of link' was either positive, negative, neutral or web convention. 'Placement of link' were: part of a list, embedded and sidebar/menu/other types.

The main category for target page was page type. The subcategories were the same as those for source pages.

2.8.3 Summary

As for citation motivation analyses, link motivations have been investigated in various circumstances without a standardised approach or a certain result. Interviews directly with authors and questionnaire surveys have been applied to find out why authors create citations. The approach is more straightforward than citation content/context analysis. However, it is not possible to interview authors or conduct questionnaire survey in link motivation analyses, because of the difficulties involved in identifying authors of web pages. As a result, link content/context analyses have been conducted extensively through various classification schemes.

The analyse of target pages has the potential to disclose which types of target pages are of interest in a link data set, while that of source pages has the potential to disclose which types of sources pages are creating hyperlinks. Link context analyses are conducted through visiting both source and target pages to find out why the links exist. In comparison, Bar-Ilan's approach is the most comprehensive one, which covers source, target page types together with link context. Harries et al.'s approach is the most complete one for analysing link context, which is from three different aspects. However, the two approaches are too time-consuming to be easily conducted in large-scale studies.

2.9 Techniques Used for Identifying Link Patterns

A number of techniques have been adapted from bibliometrics to identify link patterns in webometrics. This section discusses some of the techniques in detail.

2.9.1 Distribution of Link and Webpage Counts

Mathematical regularities have been identified in bibliometrics with regard to the distribution of author productivities; scattering of literature in a field and word frequency in a lengthy document. Three well-known bibliometric laws over formal publications are:

1. Power law. Lotka (1926) is the pioneer scientist who tried to find regularity in publication activities. Lotka's law is an 'inverse square law' of scientific productivity, which states that the number of authors with n publications is proportional to n^2. Many bibliometricians identified similar power laws with different power values, although it was first claimed to be square. The 'Matthew Effect' (Merton, 1968a), or 'rich get richer' are reflections of power laws.
2. Bradford's law (1934). Bradford's law is about the scattering of literature in a field over different journals. Divide all journals in a given field into three parts in decreasing order with regard to paper densities in that field, and suppose each part contains the same number of relevant papers in that field. If the core part contains journals as one unit, the numbers of journals in the second and third parts are n and n^2 respectively. Bradford law is very useful for librarians to determine the core of journals in any given field.
3. Zipf's law (1949). This is about word frequency in lengthy texts. If the words occurring in a lengthy text are ranked in decreasing order, the rank of a word in that text multiplied by its frequency will be a constant.

Similar mathematical regularities have been identified on the Web. With regard to power laws, Rousseau (1997) found that the domain indicators of the matching pages from a search engine followed Lotka's Law. Pennock et al. (2002) found out that 'As a whole, the World Wide Web displays a striking "rich get richer" behaviour, with a relatively small number of sites receiving a disproportionately large share of hyperlink references and traffic.' The frequencies of inlinks, outlinks and connected component sizes in academic web areas also follow power laws (Thelwall & Wilkinson, 2003a).

With regard to the Bradford's law, there is no real equivalence but the Web is made of five parts (Broder et al., 2000). They are: the Strongly Connected Component (SCC); OUT: pages which are not in the SCC but can be reached by following links from SCC; IN: pages which can reach the SCC by following links in IN, but cannot be reached from pages in the SCC; TENDRILS: are web pages connected to IN or OUT through inlinks or outlinks; DISCONNECTED: pages do not connect to any of the four parts mentioned above. The SCC is the core part of the Web. DISCONNECTED is the peripheral part. The percentage sizes of the SCC and OUT are quite similar amongst academic areas (Thelwall & Wilkinson, 2003a).

With regard to Zipf's Law, text frequencies on the Web also have been dealt with. Many high frequency words were found to be poor content indicators (Thelwall, 2005), while low frequency words are useful for clustering academic web sites along subject lines (Price & Thelwall, 2005, to appear).

Although in this study, the focus is not on identifying mathematical regularities for departmental web publication activities, the fact that similar regularities have been identified on the Web gives some confidence that it is plausible to apply bibliometric techniques to webometric data. The power law distribution of links suggests that when conducting correlation tests, the Spearman formula is more appropriate than Pearson, because power law distributions are very different to normal distributions.

2.9.2 Link Interactions with Regard to Different Top Level Domains

Leydesdorff & Wouters (1999) suggested that Triple Helix configurations on the Internet can be searched by using hyperlinks between industrial (www.*.com), academic (www.*.edu), and governmental (www.*.gov) texts. A corresponding study in bibliometrics, is that of citations between patent and scientific literature to reveal interactions between technology and science. The suggestion was firstly made by Ellis et al. (1978) and the view is supported by some pioneering work (Carpenter et al., 1980; Carpenter & Narin, 1983; Narin & Noma, 1985; van Vianen et al., 1990). In this section, the first part will briefly reviews the studies on citations between patent and scientific literatures, while the second part will review the studies on links counting to or from different web areas.

2.9.2.1 Academic-Industry Collaboration through Patent and Scientific Literature Citation Analysis

There are three different approaches to identify the relations between science and technology through bibliometric techniques. They are to investigate

1. Citations from patents to scientific literatures
2. Citations from scientific literature to patents
3. Author and inventor relations

The first would identify the influence of science on technology, while the second shows the influence from technology to science (Oppenheim, 2000). The third is through patent and scientific paper co-authorship analysis to disclose the relationship between the two. In comparison, the third approach is the most promising one. However, the third can only be conducted in small-scale analyses, since the technical difficulties in identifying inventors as authors of scientific literatures on a large scale (Coward & Franklin, 1989; Noyons et al., 1994). Amongst the three, the first has been applied the most frequently, as the links established through patent citations to scientific literature are considered more important than otherwise (Glänzel, 2003). This section mainly reviews this approach. Nevertheless, the citations from scientific literature to patents have also been investigated to disclose some interesting relationships between science and technology. For example, Glänzel & Meyer (2003) studied journal papers indexed by ISI from 1996 to 2000, which contain citations to patents from the data supplied by the U.S. Patent and Trademark Office (USPTO), and found that chemistry related subfields tended to cite patents more than other scientific areas. Among technological sectors, chemical clearly dominates, followed by drugs and medical patents as the most frequently cited categories.

The second approach has not been applied as much as the first one. The idea underlying the first approach is that patents are regarded as a reflection of technology, and the scientific citations in patents are regarded as reflections of the scientific knowledge incorporated in the corresponding invention. The results from this approach can thus disclose the relationship existing between science and technology in a particular field.

van Vianen et al. (1990) investigated the patent citations in chemical technology using the patent database from the USPTO, and found that most of the articles cited were basic scientific research. Narin et al. (1997) found that 73% of the scientific papers cited on the front pages of U.S. industry patents came from public science, which are from academic, governmental and other public institutions, while only 27% are from industrial scientists. This implies that public science has a direct, massive impact on industrial technology. They also found out that the impact is increasing rapidly, which have tripled over a six-year period from 1987-1988 to 1993-1994. In another study, McMillan et al. (2000) found that biotechnology firms rely on science to a much greater extent than large, diversified pharmaceutical companies. Tijssen (2001) studied citations to Dutch-authored research papers on USPTO patents granted during the period 1987-1996. He found not only a marked increase of patent citations to Dutch research papers but also significant differences between domestic and foreign citation patterns where domestic citation links are dominated by author-inventor self-citations and patents originating from large multinational firms. Verbeek et al. (2002) used patent citation data to link science to technology. They were able to identify those fields of technology that are highly science-interactive from those that are not. Coronado et al. (2004) provides deeper insight into the role played by science in driving the technological development of Andalucia, one of the less-favored region of the EU. Acosta & Coronado (2003) analyzed scientific citations in patents to identify the science-technology flows in Spanish regions.

The science linkages have disciplinary or national differences (Grupp & Schnoring, 1992; Schmoch, 1993; Narin & Olivastro, 1998). The reasons for these differences are caused by the way examiners, applicants or inventors incorporate Non Patent Citations (NPC) with various motivations and also with different intensities or frequencies. NPC occur most heavily in pharmaceutical, chemical and electronics patents; for pharmaceutical patents, most citations are from core basic research journals, whilst in computing, communication and transportation most citations are to applied and engineering sources rather than basic research (Narin & Olivastro, 1992). Iversen (2000) found that amongst the Norwegian patents, pharmaceuticals and instruments associated the most with scientific literature. Grupp and Schmoch (1992) and Meyer-Krahmer & Schmoch (1998) found that in their studies the closest links with science are in the field of biotechnology, together with other areas related to chemicals production and information technology. The patents in their studies were from European Patent Office (EPO). The linkages also have national differences, such as Japanese patents emphasis on electronic, while those from the U.S. and U.K. have strengths in pharmaceuticals (Narin & Olivastro, 1992). The examiner citations in most countries are different from those in the U.S. (Kaback et al., 1994).

There are some limitations for the patent citation studies. Firstly, only partial science and technology interaction can be identified, since other interactions may not appear as patent citations, such as academic and industry collaborations without scientific citations in the invented patents. Secondly, the patent databases for extracting patent documents are different, and the differences may affect the results. For example, the frequencies of citations in patents from USPTO are higher than those from EPO patents (Acosta & Coronado, 2003). Narin & Olivastro (1998) also found that there was a steady increase of scientific literature citations in patents from USPTO rather than those from EPO, and for two very similar patents from USPTO and EPO, the citation overlap is very tiny. The variations of patent citations from different databases were identified by Simmons (1994) in another study. Thirdly the patent citation analysis itself lack of validation.

Generally speaking, the studies in aggregate seem to confirm that patent citations to scientific literature measure the intensity of the science supporting technology (Narin, 1994; Tijssen et al., 2000; Tijssen, 2001). However, there are some disputations. Schmoch (1993) studied patent citations in the fields of lasers and polyamides, and concluded that the using of citation data to link science and technology was not justified. Vinkler (1994) studied 250 pharmaceutical patents from different countries, and concluded that the examiner citations had no validity, only applicant citations could lead to valuable inference about the science and technology links. The most detailed criticism of patent citation analysis was made by Kaback et al. (1994). They emphasized the differences between journal citations and patent citations. Unlike the authors of journal papers, those of patent applications are more likely to cite unsuccessful approaches to the question they solved, in order to claim more monopoly right. Meyer (2000) studied patent citations in the field of nanoscale technologies and found that the citation linkages hardly represent a direct link between cited paper and citing patent. In order to fully validate patent citation analysis, Oppenheim (2000) suggested ten research questions. These include examiner and applicant's citation motivation analysis; several correlation tests with existing non patent data; the quality of searches carried out by examiners and the overlap between examiner citations and applicant citations. Although patent citation analysis lack a complete validation, it is a useful information source to disclose some interesting interaction patterns between science and technology.

2.9.2.2 Online Interactions between Different Web Areas

In comparison, the interactions amongst different organizations on the Web might be revealed by counting links to or from different top-level domains. For example, gov domain mainly represents government web sites, while com domain includes many commercial web sites, hence the links between gov and com domains have the potential to disclose online interactions between web sites from government and industry organizations.

Thelwall (2001a) studied link interaction between the seven general Top Level Domains (gTLD) (ICANN/GNSO, 2001), and also those between a heterogeneous selection of large web sites (a selection of British universities, large computing companies and web directories). He used the search engine AltaVista to count the links between various web areas. The seven gTLDs are:

- edu: domain for educational web sites mainly from the USA and Canada
- mil: domain for the United States Military
- net: domain mainly for network organizations, but open to all
- gov: domain for the United States Government
- com: domain mainly for commercial organisations, dominated by North America, but open to all
- int: domain for organizations established by international treaties between governments or Internet infrastructure databases
- org: domain mainly for organizations not belonging to the above mentioned domains and also not belonging to a country domain, but open to all

He found that the com domain was the dominant source of external links for other top level domains. Amongst the set of large web sites, business relationships were identified through the links between them. The most apparent connections were found between Microsoft and MSN, and between Netscape and Excite. In addition to this, other interesting links have been identified between Cambridge University and York University; between Cambridge University and Sun company; between York University and Warwick University; between Sun company and Oracle company.

There is a set of studies that identified the academic interactions on the Web between different countries (Smith & Thelwall, 2002; Thelwall & Smith, 2002; Thelwall et al., 2003b). Smith & Thelwall (2002) counted links between Australia, New Zealand and the UK academic domains using search engine AltaVista. They discovered that Australia and the UK interconnect the most, while New Zealand connect to Australia more than with the UK; New Zealand was found to be isolated on the Web relatively to two other countries. The same authors studied links amongst 13 Asia-Pacific countries' academic domains (Thelwall & Smith, 2002). They found that Australia and Japan were clearly the heart of the Web in the region. In another study, Thelwall et al (2003b) studied the academic interlinking amongst 14 Western European countries to identify linguistic link patterns. Again search engine AltaVista was used to collect relevant link data. However, not every European country has an official academic domain, thus each country's university domain list was used to represent its academic domain for link counting purposes. They found that international interlinking throughout Europe in English, and additionally in Swedish in Scandinavia; linking between countries share a common language; Greece was found to be isolated in that web area.

In addition to count links between various web areas, word invocations have been employed to identify online interactions. Thelwall (2004a) tried to find out whether it is possible to identify university-industry knowledge transfer on the Web. He analysed the invocations of pure and applied science journals in the Web, focusing on commercial sites. In that study, he searched the invocation of relevant journal names through the search engine Google, and then categorized those result pages into academic or non-academic ones by identifying whether their urls has '.edu' or '.ac.' strings. He classified the non-academic web pages to identify the reasons why those journal names were invoked. He found evidence that applied research was more highly invoked on the non-academic web than pure research. However, on a micro level, fewer evidences show the transfer of academic knowledge to a commercial setting. He then suggested that it might be for commercial reasons that companies are reluctant to put online the research they benefit from to avoid giving intelligence to their competitors. Thus he concluded the Web is not a suitable source to yield useful information about university-industry knowledge transfer.

Leydesdorff & Curran (2000) mapped and compared university-industry-government relations on the Internet for Netherlands and Brazil. They used AltaVista's advanced search facilities to limit web pages within each of the two country domains, and search for the invocations of various text combinations of 'university', 'industry' and 'government'. Although they did not count links between different domains as suggested by Leydesdorff & Wouters, they found similar patterns for the two countries. 'industry-government' lag behind when viewed from international perspective, while 'university-industry' lag behind when viewed from national perspective.

It was once possible to count links from different geographical regions through AltaVista (Tang & Thelwall, 2004). Interesting regional link patterns were identified in that study: apart from Canada, the US history, chemistry and psychology departments were found closely connected with Europe and Asia rather than the rest of the American continent. Unfortunately, the advanced facility of AltaVista to limit the source regions ("location: by regions") was removed soon after that study.

There are some webometric studies counting links from different domains to academic web sites (Thelwall, 2002a; Li et al., 2003b). Thelwall (2002a) counted inlinks to a set of UK universities from the seven gTLDs and other relevant national domains, but the purpose was to identify a better source for counting research related links to universities and avoiding labour extensive classification of web pages. The assumption was that inlinks from academic domains in other countries might represent international recognition, since international links show more intention to link than national ones although not necessarily with more technical effort. Better source of inlinks were not identified in that study, since the edu, external, ac.uk, uk and org WIFs correlate similarly with the research averages. Nevertheless, this opened the door to disclose interactions between universities' web sites and other national or international web areas. This is an analogy of citation analysis between patents and scientific literature to illustrate interactions between science and technology.

Apart from universities, links to departments from different domains were counted in the preliminary study as described in section 3.1 (Li et al., 2003b). In that study, the most links are from – external, ac.uk, co.uk, uk, edu, org, com, and net domains, with a sparse number of links from mil, int and gov domains. This may imply that departmental web sites simply attract less inlinks from military and government organizations on the Web. The aim in that study was to test whether inlinks from various web areas correlate significantly with research performances rather than to identify link interactions between departments and various web areas. However, no known studies have investigated outlinks from academic web sites to different domains. With the results, a full picture can be illustrated on how the web sites interact with various web areas. Compared with patent citation data collection, the link interaction data to or from different web areas is relatively easier and cheaper, although the quality may not be at the same level.

There are limitations for identifying online interactions through investigating links to or from various domains. Firstly, domains may not guarantee the types of web sites within them. For example, even if web sites within com domain are mainly commercial related, it may contain personal web sites. This is similar as to identify academic-industry collaboration through patent citation analysis, where patents are created mainly by industrial organizations but also by universities (Henderson et al., 1998). Secondly, as discussed in section 2.3 search engines only cover the Web partially. The results thus can be used as indicative rather than definitive. Just like patent citation analysis, we cannot rely on it, but if we ignore it, we simply lose some interesting information.

2.9.3 Link Weighting

A fundamental issue for link counting is that currently all links have been regarded as the same in webometric studies. A link made by a first year student may not be as important as one created by a professor. The idea to give counts different weights is not new. In bibliometrics, it is more sensible to give a higher weight to a citation from a prestigious journal than a trivial one (Pinski & Narin, 1976); publication counts are weighted according to the status of the journals in which they appear (Martin, 1996), or the number of written pages in them (Luwel et al., 1999).

In computer science, link-based ranking algorithms have been widely used by commercial search engines. For example, the heart of Google software is PageRank which gives each web page weight to pass on to the other pages through links (Brin & Page, 1998; 2002). Another link-based algorithm for search engines is Hyperlink Induced Topic Search (HITS) (Kleinberg, 1999). HITS used inlink weights to determine a web page's authority score, and the weight of outlinks to determine a web page's hub score. HITS is designed to identify web pages that are most useful to a user's query, while PageRank is applied to the whole Web regardless of users' queries.

PageRank is a natural way to rate pages, and the success of Google confirms the efficiency of the algorithm. Nevertheless, it includes links within a web site (site selflinks) and hence it is not a good technique for webometric purposes. For example, PageRank failed to identify the most important web page for the universities in the UK, Australia and New Zealand due to the domination of internal links (Thelwall, 2003a). At present, no existing technique is efficient enough to weight links according to their importance in webometrics. This is one of the reasons why link metrics are not suitable to measure individual units' web performances.

2.9.4. Link Propensity

In bibliometrics, the strength of co-authroship links has been measured through Salton's meausure: the number of joint papers for a pair of countries divided by the square root of the product of the two countries total publications (one country's publications multiplied by the other's) (Glänzel & Schubert, 2001). Through Salton's measure of co-authorship, many interesting patterns have been identified. For example, Australia connects highly to the UK, while Canada connects highly to the USA (Glänzel, 2001; 2001). However, the results may not be reliable if the differences of publication sizes amongst a set of countries are too large.

In webometrics, Link Propensity (LP) has been used to illustrate the tendency of two web sites to link to one another (Smith & Thelwall, 2002). By definition, LP is the number of external inlinks to a university web site or collection of such sites, divided by the product of the number of academic staff members from both the source and target universities (sometimes the number of web pages are used instead if numbers of staff members are not available). The LP factors out the size effects from source and target sites. LP helps to identify many interesting interlinking patterns. These are:

- The com domain was found to be a major source of external links amongst the seven gTLDs (Thelwall, 2001a);
- Universities in New Zealand were found to be relatively isolated on the Web, compared with those in the UK and Australia (Smith & Thelwall, 2002);
- Amongst Asia-pacific universities, Australian and New Zealand are found to interlink heavily (Thelwall & Smith, 2002);
- English is the dominant language in academic web pages in Western European countries, and countries tend to link to each other more if they share the same languages (Thelwall et al., 2003b).

LP is different from Salton's measure of co-authorship. LP can measure the tendencies for two web sites to link to each other from both directions, while Salton's measure can only give one value for each collaboration pair. The limitation for LP, however, is that it can only measure the link propensities between two units. When comparing the ability for a set of departments to attract inlinks from their peers in two other countries, a new technique is needed.

2.9.5 Co-citation and Bibliographic Coupling on the Web

Co-citation and bibliographic coupling are techniques applied in bibliometrics to identify clusters of same or related research topics. Co-citation analyses papers that are cited by same papers (Small, 1973), while bibliographic coupling is about those citing papers that share same references (Kessler, 1963). There is much literature about co-citation analysis in different contexts.

- Author co-citation analysis in Information Science (McCain, 1986; Bayer et al., 1990; White & McCain, 1998);
- Journal co-citation analysis in Economics (McCain, 1991);
- Identifying pathways that transverse different disciplines through co-citation analysis (Small, 1999a; Small, 2000);
- Constructing a map of science through co-citation clusters (Small, 1999b);
- ISI Atlas of Science based on paper co-citation analysis (Garfield, 1987)

Bibliographic coupling has been applied together with co-citations, and the two have been found to complement each other (Sharabchiev, 1988). ISI SciViz is the product of the combination of the two techniques (Small, 1998).

Based on the structural similarities between hyperlinks and citations as discussed in section 2.2.1, co-citation and bibliographic coupling techniques have been applied to the Web for identifying clusters of related topics. In webometrics, co-linked web pages are analogous to co-citations, while co-linking web pages are analogous to bibliographic coupling (Björneborn, 2001b). Similar topics on the Web have been identified through analysing co-linked web pages (Larson, 1996). The combination of links, co-linked and co-linking web pages gives the highest probability of identifying similar web sites, while co-linked or co-linking web pages analysis on its own is not so efficient (Thelwall & Wilkinson, 2004).

To conduct co-linked or co-linking web pages analysis, the link data must be large enough. Since the number of inter-departmental links is sparse (Harries et al., 2004), it may not be plausible to apply the two techniques for departmental interlinking.

2. 10 Summary

The structural similarity between citations and hyperlinks is the rationale for webometric studies to apply techniques from bibliometrics to the Web. With regard to bibliometrics laws, similar mathematical regularities have been identified on the Web. Co-linked and co-linking techniques are based on co-citation and bibliographic coupling from bibliometrics. WIF is the application of JIF on the Web. However, citations and hyperlinks are from different environment. Formal scholarly communication patterns can be identified through citation analysis, whereas the results from link analysis reflect informal scholarly communication.

Both search engines and a personal web crawler SocSciBot have been used to collect link data. Search engines are good at covering large web areas, but they can only cover the Web partially and their search algorithms are opaque from users. Personal web crawlers, on the other hand, can cover separate web sites extensively, although they cannot cover large web areas. Using SocSciBot, links can be counted at different ADMs level by manipulating urls of web pages. The rationale for doing this is to remove anomalies at higher aggregated document levels.

Statistical correlation tests between link metrics and research measures, together with the identification of link motivations can be used to validate link data. Citation counts are regarded to be the most relevant data to be compared with inlinks. Other indicators such as peer review evaluations and research grants can also be used to further validate link data from different perspectives.

Just as patent citation analysis has the potential to disclose science and technology interactions, links to or from various web domains have the potential to disclose online interactions between academic web sites and other organizations.

3. Preliminary Research

This chapter reports two preliminary studies which help to gain insights to design and execute the main experiments of this study. The first study was designed to investigate whether there are any significant correlations between link and research measures for UK computer science departments (Li et al., 2003b). The second study was designed to ascertain how important it is for academic web pages to be in English, in order to attract inlinks from other universities in Western European countries (Li et al., 2003a).

3.1 An Investigation into Computer Science Departmental Interlinking in the UK

The purpose of this study was to assess whether a webometric analysis of departmental level web site interlinking was practical, and whether there was a connection between research and links at this level. A secondary aim for this study was to identify link anomalies in more details at the departmental level. Computer science departments were chosen, as they were hypothesised to make more use of the Web than most other departments, and so would be most likely to host enough links to give significant results.

3.1.1 Methods Adopted

3.1.1.1 Identifying the Domain Names for the Computer Science Departments

The domain or directory name list of the 79 computer science departments that submitted in the 2001 Research Assessment Exercise (RAE) (HERO, 2001d) were identified manually. Firstly, the universities' homepages were identified through Google, then by following relevant links the domains or directories of computer science departments within the universities were identified. Secondly, the 'UK Computer Science Department' link list hosted in the University of Wales Swansea's web site (http://www.swan.ac.uk/compsci/resources/UKCSDepts.html) was cross checked with those found manually. Relevant domains or directories, which were different from those already identified and still work, were added to the domain and directory list. Thirdly, the syntax:

host:xxx.ac.uk AND link:www.cs.*.edu

was used in AltaVista (2001) with the hypothesis that the computer science departments in the UK are quite likely to make links to computer science departments in the U.S.. More old domains or directories that still work were identified, and added to the list.

In the meantime, 79 e-mails were sent to computer science departments' webmasters (or a staff member instead if the webmaster was not available) and 31 replies were received. Although most of them only confirmed their current domains or directories, some gave the urls of research groups belonging to their departments, which otherwise could be difficult to identify as they might not be in the hierarchy of their department's domain. Finally, the derived domain or directory name list was compared with the domain list of the 109 UK universities available from the web site of the Statistical Cybermetrics Research Group in the

University of Wolverhampton (2001), and more domains or directories were identified and added to the list.

3.1.1.2 Data Collection for the 79 Computer Science Departments

1. Link data collected from AltaVista

The number of inlinks from 10 different domains (org, edu, uk, ac.uk, co.uk, com, mil, net, gov, int) and external inlinks (from the rest of the Web) to each department was collected through AltaVista's advanced query facility. In early research, identical Boolean queries in AltaVista have returned different results, and researchers have had to develop methods to mitigate this problem (Ingwersen, 1998; Smith, 1999c; Smith, 1999b; Smith, 1999a). Fortunately, AltaVista has become more stable over time (Thelwall, 2001d; Thelwall, 2001b; Vaughan & Thelwall, 2003). It should be noted, however, that the data collected by any search engine is inherently incomplete (Lawrence & Giles, 1999), and the results should therefore be interpreted with caution. In the UK, a lot of universities have more than one version of their domain name, although this may not be the case for the rest of the world. External inlinks are defined to be inlinks to the department that do not come from the same university. The external inlinks to the division of informatics (the whole division was submitted to 2001 RAE as computer science) in the University of Edinburgh can be obtained through the following syntax:

(link:.dcs.ed.ac.uk OR link:.dcs.edinburgh.ac.uk OR link:.informatics.ed.ac.uk OR link:.cogsci.ed.ac.uk OR link:.cogsci.edinburgh.ac.uk OR link:.hcrc.ed.ac.uk OR link:.aiai.ed.ac.uk OR link:.dai.ed.ac.uk) AND NOT host:.ed.ac.uk AND NOT host:edinburgh.ac.uk

The syntax to collect the links from edu domain to the division is as follows:

(link:.dcs.ed.ac.uk OR link:.dcs.edinburgh.ac.uk OR link:.informatics.ed.ac.uk OR link:.cogsci.ed.ac.uk OR link:.cogsci.edinburgh.ac.uk OR link:.hcrc.ed.ac.uk OR link:.aiai.ed.ac.uk OR link:.dai.ed.ac.uk) AND domain:.edu

The syntax to find the number of web pages in the division is as follows:

host:dcs.ed.ac.uk OR host:dcs.edinburgh.ac.uk OR host:informatics.ed.ac.uk OR host:cogsci.ed.ac.uk OR host:cogsci.edinburgh.ac.uk OR host:hcrc.ed.ac.uk OR host:aiai.ed.ac.uk OR host:dai.ed.ac.uk

Mirror sites, once identified, were eliminated. For example, mirror site http://src.doc.ic.ac.uk, which is hosted in the site of computer science department in Imperial College, but created by the Sun Microsystems, is excluded when counting links to the department. The syntax below shows how to count links from the ac.uk domain to the department, excluding the mirror site.

(link:.doc.ic.ac.uk AND NOT link:src.doc.ic.ac.uk) AND host:.ac.uk AND NOT host:.ic.ac.uk

2. Link data collected by SocSciBot
As discussed section 2.3, SocSciBot was created in response to the deficiencies in the early days of search engines (Bar-Ilan, 1999; Lawrence & Giles, 1999; Rousseau, 1999; Thelwall, 2000). The details of how SocSciBot extracts the link data from a web site, and how the relevant link information can be derived through the tools are available online (Thelwall, 2001c).

A piece of software called 'subsite extracter' and the departments' domain or directory names were used to extract their link structures from the existing 109 UK universities' link data, which was collected in July 2001 by SocSciBot (The Statistical Cybermetrics Research Group, 2001). Associated software was used to calculate the number of links both to and from the computer science departments. The number of web pages in each department is also derived in this way.

3. The number of academic staff members in each department
The number of academic staff members are derived from 'Staff FTEs by Institution and Cost Centre 1998/99' (HESA, 2002). As some computer science departments' domains or directories are shared by other disciplines, in order to calculate a fairer result, the number of academic staff members involved from other disciplines are included. For example, the domain name of the computer science department in the University of Hertfordshire-'feis.herts.ac.uk', is shared by the department of aerospace, civil and mechanical engineering (ACME); the department of electronic, communication & electrical engineering (ECEE); the department of manufacturing systems engineering (MSE); maths and music. All of the academic staff members from the other departments are included.

3.1.1.3 The Calculation of WIFs and Correlation Coefficients
Two versions of WIFs are used, one with the number of academic staff members from a department as denominator, which is represented as the WIF (staff), and another with the number of web pages in a site as its denominator. The numerators are all the number of external inlinks.

The Spearman rank order correlation coefficient rather than the Pearson correlation coefficient was used to calculate the correlations between the links and the RAE measures. This is because the frequency of inlinks forms a highly skewed distribution, for which Pearson is inappropriate. In order to simplify the calculation, the different RAE values are transferred into ordinal numbers (RAE: 1, 2, 3b, 3a, 4, 5, 5* into 1, 2, 3, 4, 5, 6, 7 respectively). The approach has been used by previous research (Thelwall, 2001b) and newspaper ranking schemes. Section 2.7.2 discusses this in detail.

3.1.2 Results

3.1.2.1 The AltaVista Results
The results from AltaVista are displayed in table 3.1. The second column shows the correlation coefficients between the numbers of external inlinks from different domains and their research productivities (RAE rating multiplied by the number of research staff members submitted). The third column shows the correlation coefficients between the WIFs (staff) and the RAE averages. Column 4 illustrates the correlation coefficients between the WIFs (with number of web pages as denominators) and the RAE averages.

The results in column 4 are not as significant as those of columns 2 and 3. When counting web pages with AltaVista, no page was found for the computer science departments in the University of Cambridge and Oxford Brookes, only one page for the University of Glamorgan and two pages for Bournemouth University. This is clearly incorrect. Apart from the fact that the number of web pages in a site is not suitable to represent the size of the relevant entity (see section 2.5.2), the count itself is unreliable. Again web pages as denominators have been confirmed to be unreliable in WIFs calculations (Thelwall, 2001c, 2002a-b).

Table 3.1 Spearman correlation coefficients between link and research measures. (* = significant at the 5% level, ** = significant at the 1% level, ***=significant at the 0.1% level)

Source domain	Between inlinks and research productivities n=79	Between WIFs(staff) and RAE averages n=79	Between WIFs and RAE averages n=77
org	0.645**	0.674**	0.349**
edu	0.606**	0.609**	0.275*
uk	0.603**	0.559**	0.119
External	0.600**	0.604**	0.256*
ac.uk	0.599**	0.566**	0.092
com	0.592**	0.586**	0.178
mil	0.582**	0.551**	0.399**
co.uk	0.556**	0.489**	0.107
net	0.553**	0.561**	0.311**
gov	0.529**	0.512**	0.304**
int	0.290**	0.254*	0.179

Most of the inlinks to the departments are from the domains: ac.uk, uk, edu, org, com, co.uk, and net. Figures 3.1 to 3.6 show the relationship between link and research measures. The outliers in these diagrams are as follows.

Outliers in Figure 3.1

- Imperial College (298.2, 3233)
- University of Cambridge (240.1, 3166)
- University of Oxford (252, 2742)
- University of Strathclyde (53.2, 606)
- De Montfort University (42.5, 502)
- University of Brighton (25, 416)
- University of Wolverhampton (17.4, 254)

Outlier in Figure 3.2

- Imperial College (298.2, 62921)
- Napier University (32, 10496)
- University of Wolverhampton (17.4, 5787)

Outliers in Figure 3.3

- Imperial College (298.2, 16514)
- University of Wolverhampton (17.4, 1703)

Outliers in Figure 3.4

- University of Cambridge (7 53.39)
- University of Oxford (6, 50.87)
- University of Stirling (3, 8.17)

Outliers in Figure 3.5

- University of Cambridge (7, 547.55)
- University of Oxford (6, 442.04)
- University of Aberdeen (5, 181.43)
- Napier University (4, 156.67)
- University of Stirling (3, 113.56)
- University of Wolverhampton (2, 74.67)

Outliers in Figure 3.6

- Imperial College (7, 95.24)
- University of St Andrews (6, 78.42)
- University of Stirling (3, 15.13)
- University of Wolverhampton (2, 21.97)

The School of Computing and Information Technology in University of Wolverhampton was observed as an outlier in Figures 3.1 to 3.3, 3.5 and 3.6. 'UK Sensitive Map' (clickable maps for academic organizations in the UK) (http://www.scit.wlv.ac.uk/ukinfo/uk.map.html) is hosted in the site, and attracts a lot of links to the department. Table 3.2 illustrates the link counts from different web areas to the clickable map. This also shows the fact that AltaVista's results are now (normally, but not always) logically consistent. The top figure should logically be the sum of the other two, and it is.

Table 3.2 The AltaVista link counts to the clickable map in the University of Wolverhampton

Syntax	result
link:.scit.wlv.ac.uk/ukinfouk.map.html AND host:.ac.uk	1054
link:.scit.wlv.ac.uk/ukinfouk.map.html AND host:.ac.uk AND NOT host:.wlv.ac.uk	1030
link:.scit.wlv.ac.uk/ukinfouk.map.html AND host:.wlv.ac.uk	24

The department of computing in Imperial College was observed as an outlier in figures 3.1 to 3.3 and 3.6. The mirror site: SunSite Archive (http://src.doc.ic.ac.uk/) was spotted, and the link pages to this site were deleted. Later, the site http://sunsite.doc.ic.ac.uk was found to be an alias of the same SunSite Archive mirror site. The number of link pages to this mirror site resulted in 16,408 (external), 606 (edu) and 1,310 (ac.uk) hits from AltaVista. If these numbers were removed, then at least in figures 3.1 and 3.2, the department would not appear as an outlier.

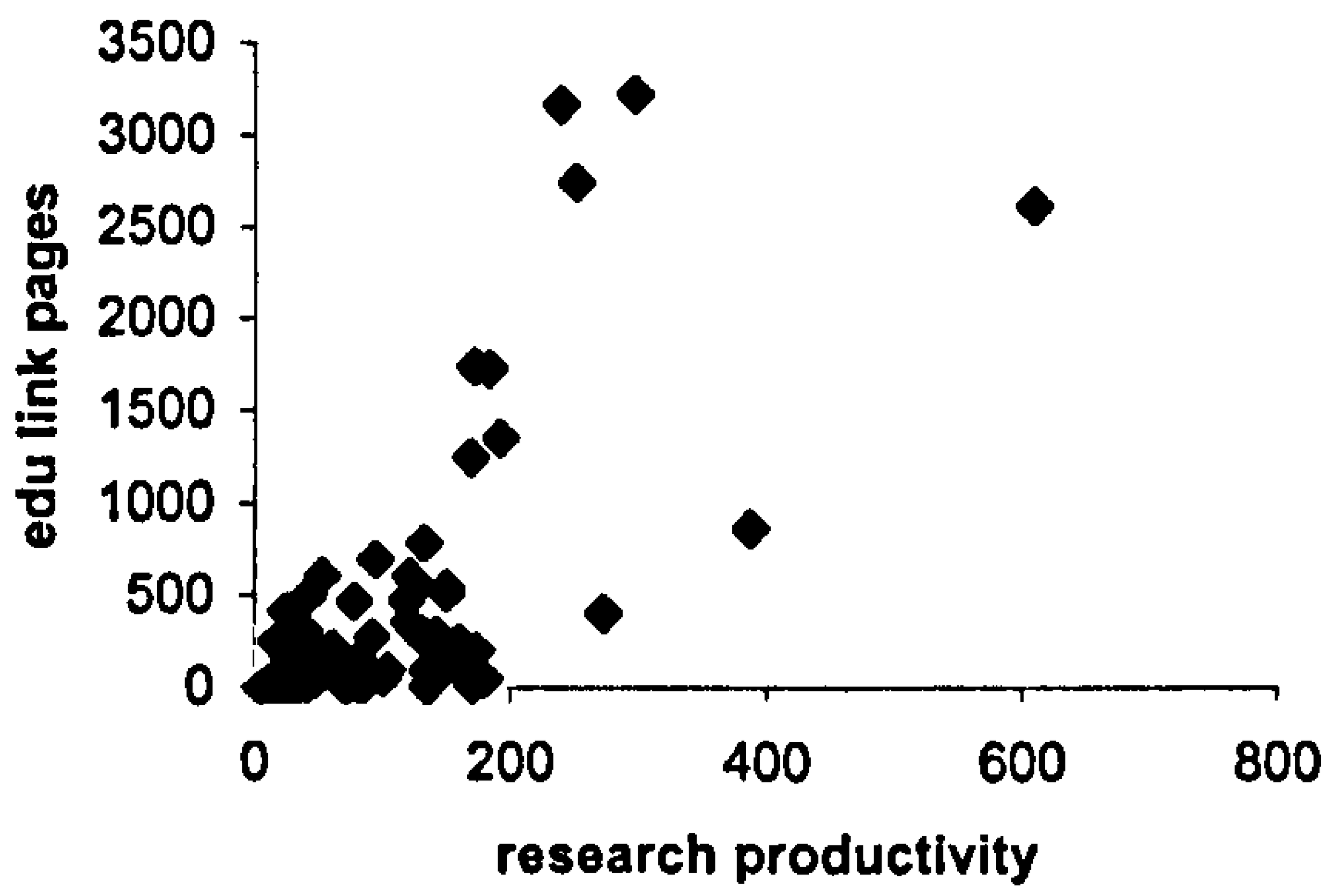


Figure 3.1 edu inlinks (from AltaVista) against research productivities for UK computer science departments

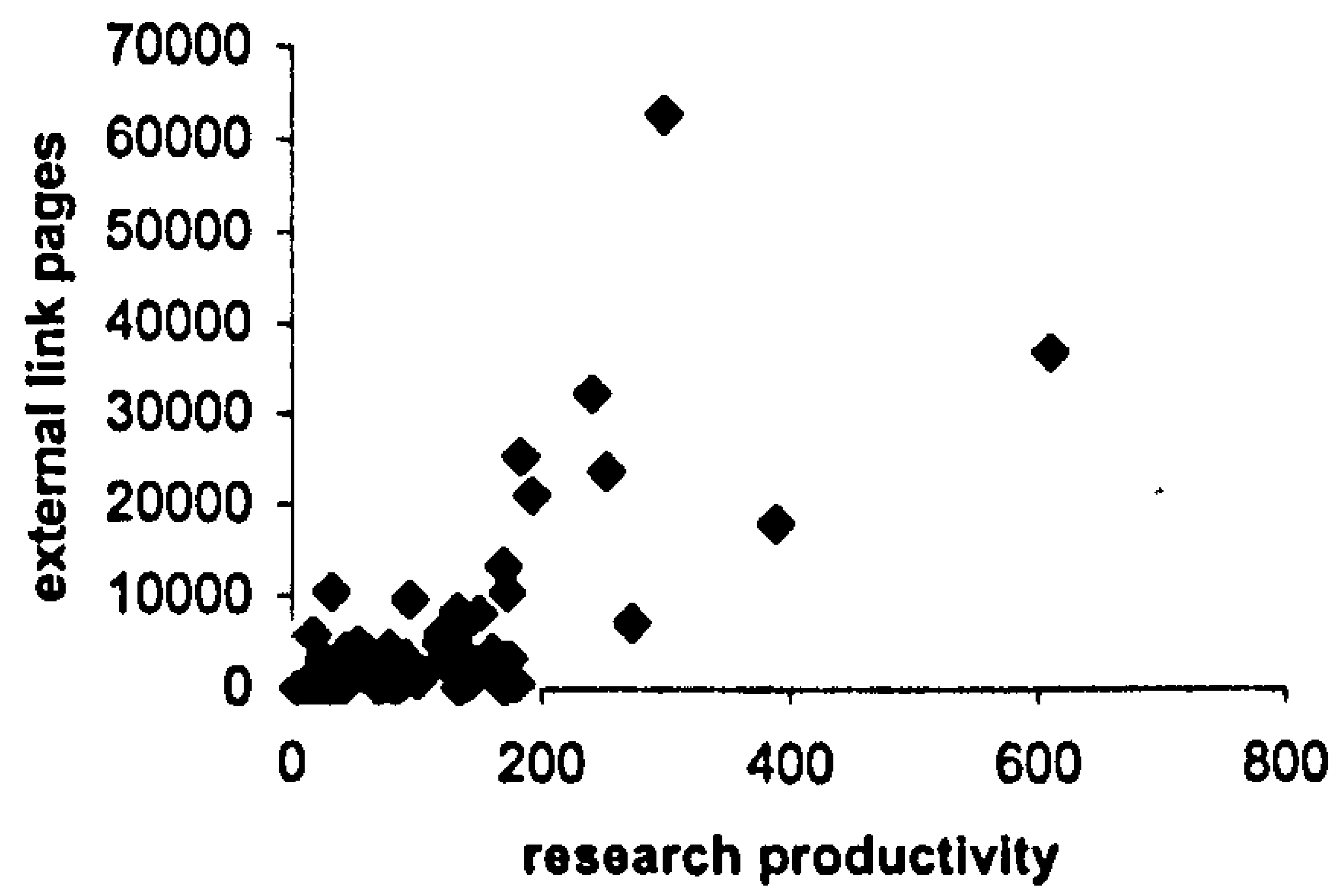


Figure 3.2 External inlinks (from AltaVista) against research productivities for UK computer science departments

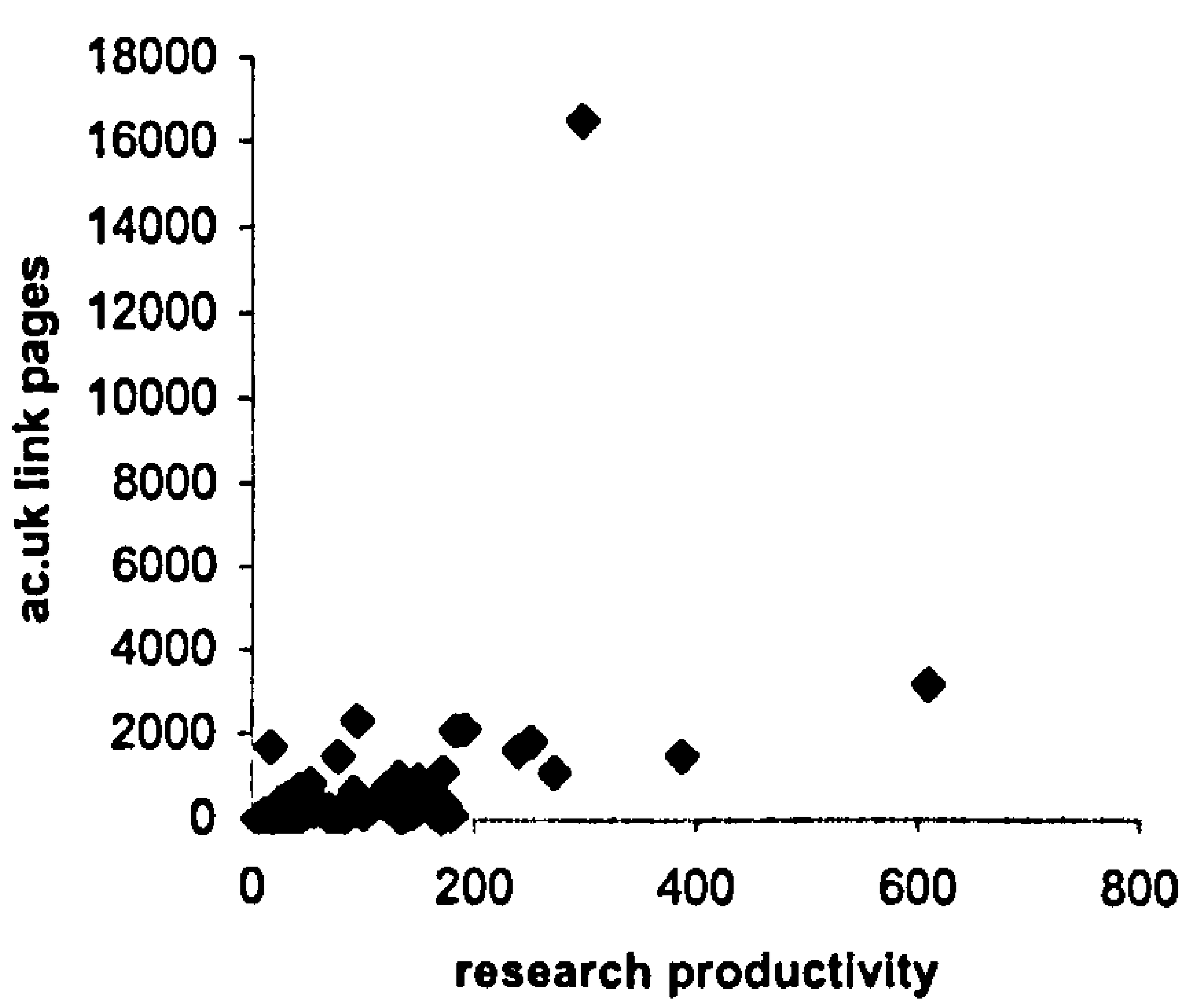


Figure 3.3 ac.uk inlinks (from AltaVista) against research productivities for UK computer science departments

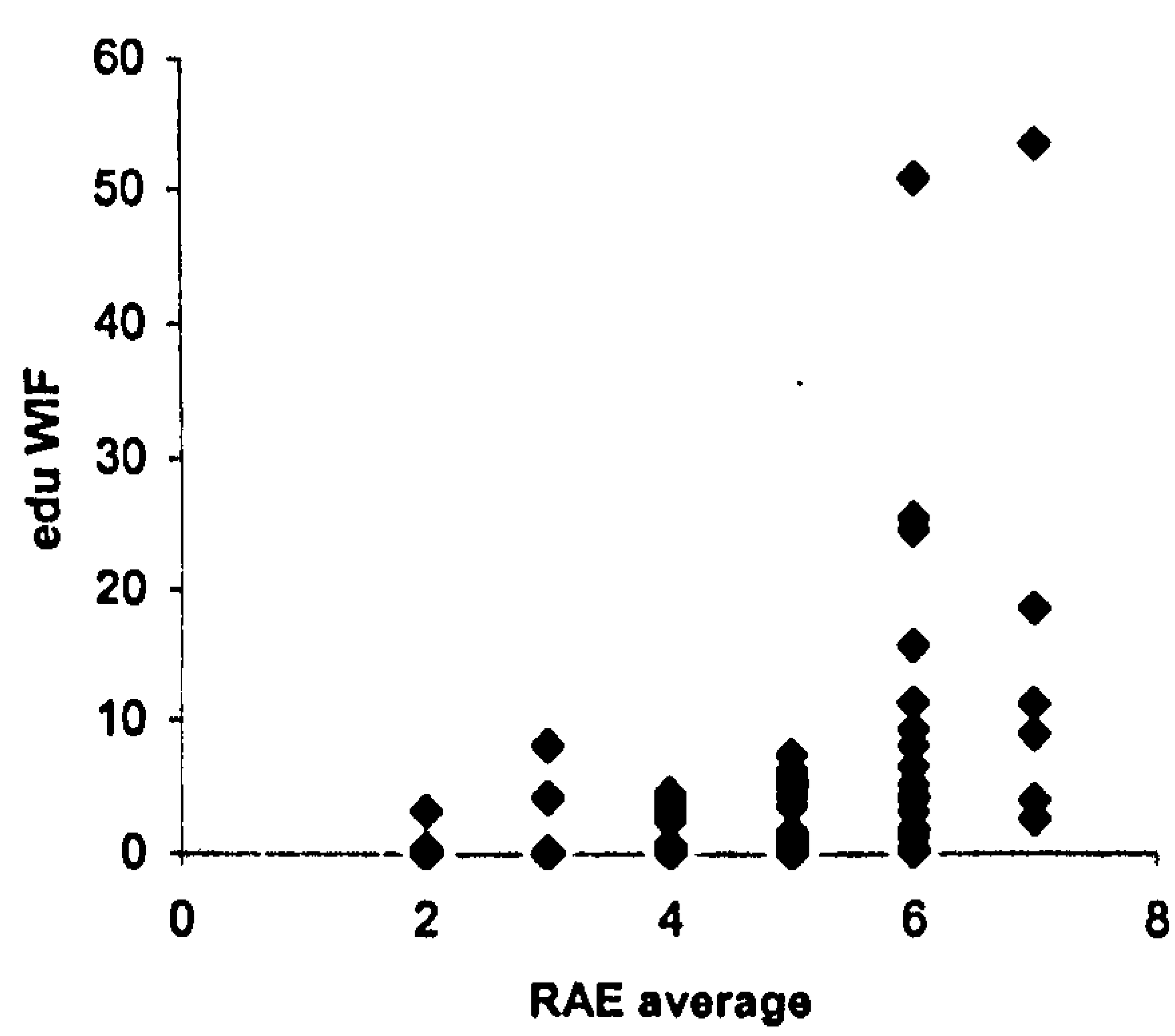


Figure 3.4 edu WIFs (staff) against RAE averages for UK computer science departments

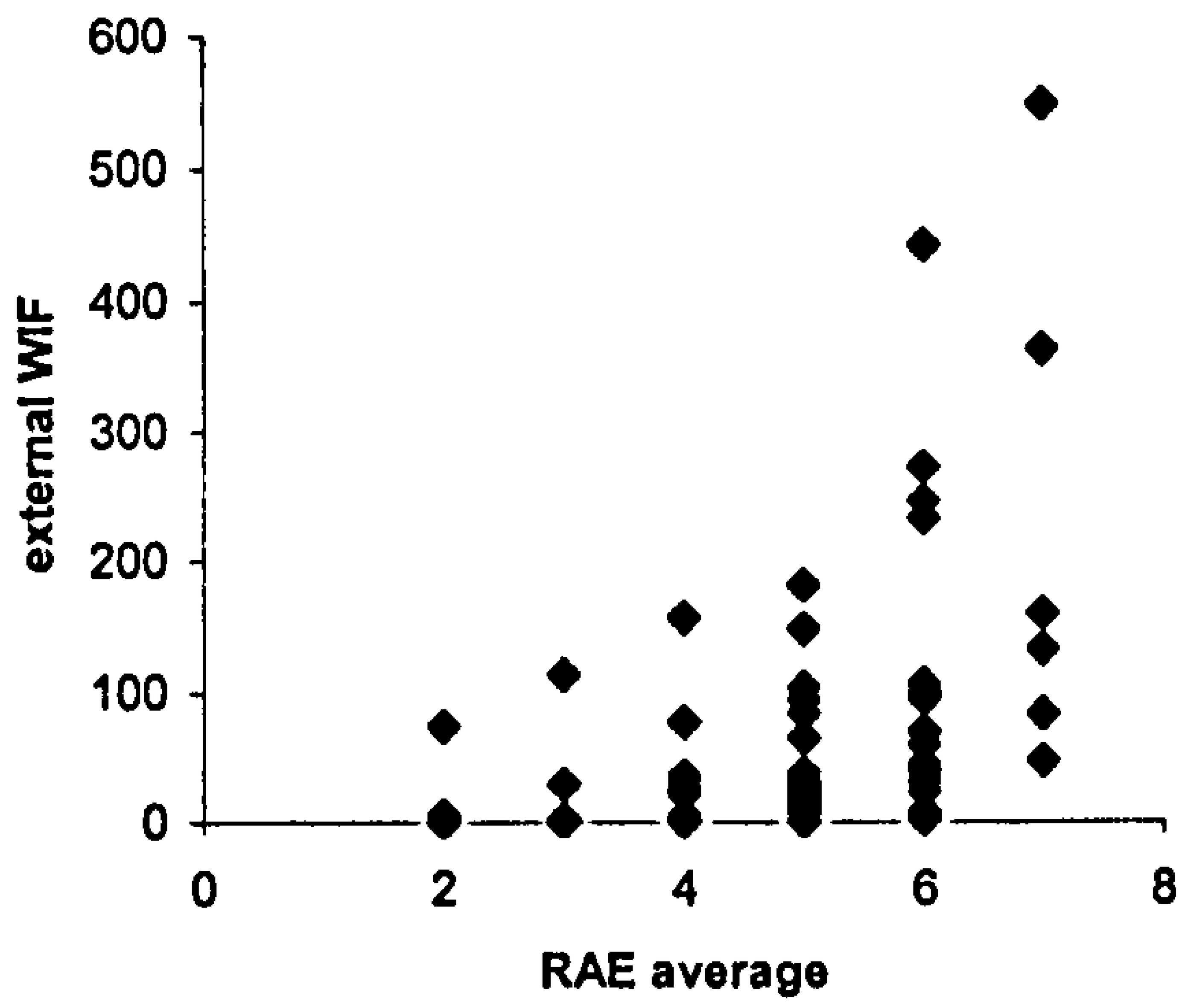


Figure 3.5 External WIFs (staff) against RAE averages for UK computer science departments

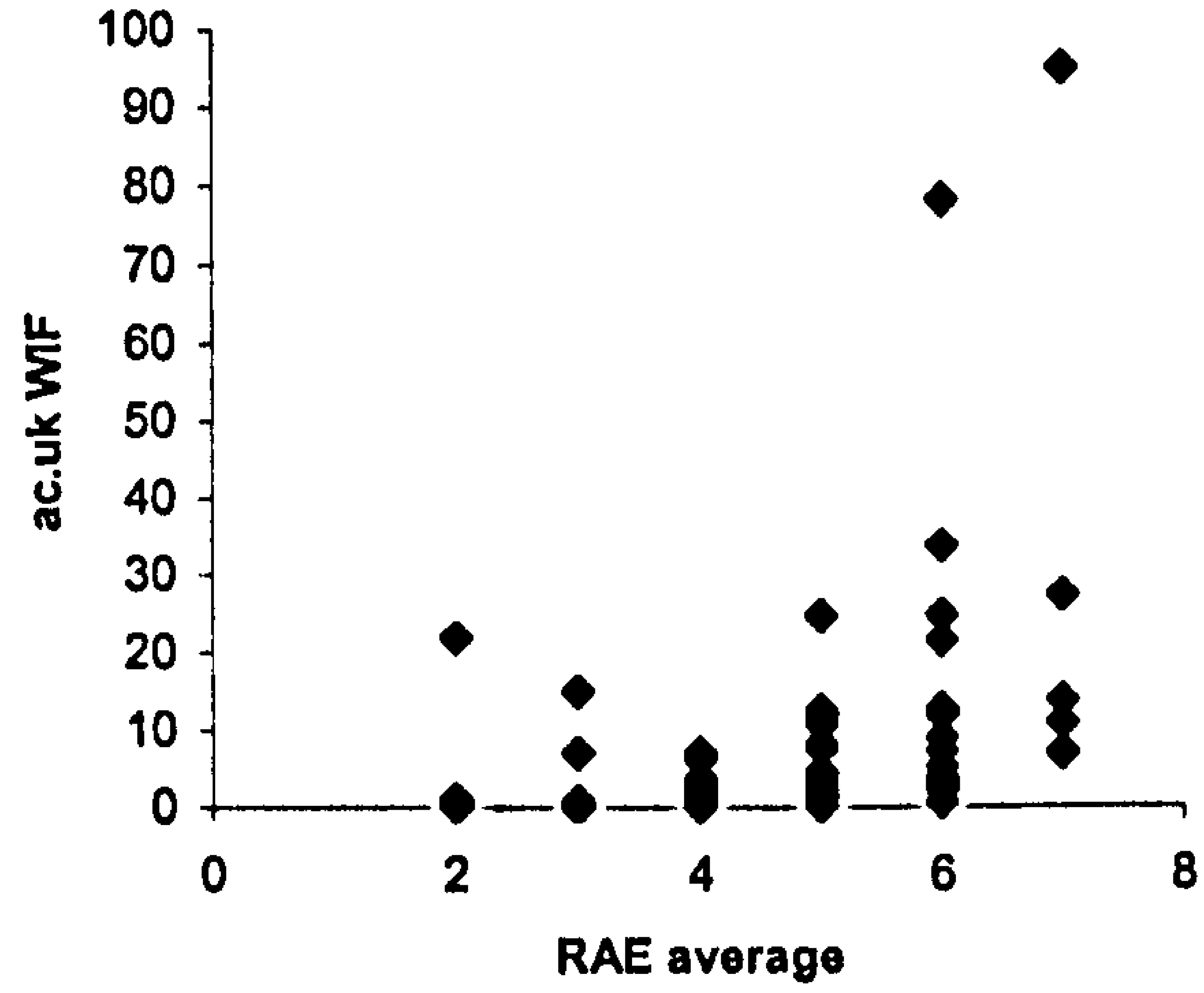


Figure 3.6 ac.uk WIFs (staff) against RAE averages for UK computer science departments

3.1.2.2 The SocSciBot results

The correlations between different versions of WIFs and RAE averages, and between the numbers of inlinks and the research productivities are reported in table 3.3. The correlation between the WIFs with web pages as denominators and RAE averages is not significant at all. However, the other two values are significant at the 1% level.

Table 3.3 Spearman correlation coefficients between link (from SocSciBot) and research measure (* = significant at 5% level, ** = significant at 1% level, n=70)

Link and research measure	Correlation coefficient value
Inlinks and research productivities	0.494**
WIFs (staff) and RAE averages	0.411**
WIFs and RAE averages	0.083

Chu et al. (2002) stated that the number of inlinks can illustrate the 'visibility' of a web site, while the number of outlinks can illustrate its 'luminosity'. Although Chu et al. (2002) were not able to find any significant correlations between inlinks and outlinks of the 53 North American library and information science departments, inlinks correlated significantly with the outlinks for the 79 UK computer science schools (Spearman correlation coefficient value is 0.601, significant at the 1% level). Figure 3.8 illustrates a strong linear trend between inlinks and outlinks for the 79 computer science departments. The most apparent outlier is Reading University (93536,1730). If the department were deleted, then the correlation coefficient value would be higher at 0.738**. Generally, the departments that link more to others are also more likely to be linked to by their peers.

Figure 3.7 shows the relationship between inlinks and research productivities for the 79 UK computer science departments. The school of computing and information technology in the University of Wolverhampton again is an outlier in figure 3.7. The reason might be the same as the previous AltaVista result.

Outliers in Figure 3.7

- University of Birmingham (132.6, 508.2)
- Heriot-Watt University (95, 4149)
- Napier University (32, 2742)
- University of Wolverhampton (17.4, 2800)

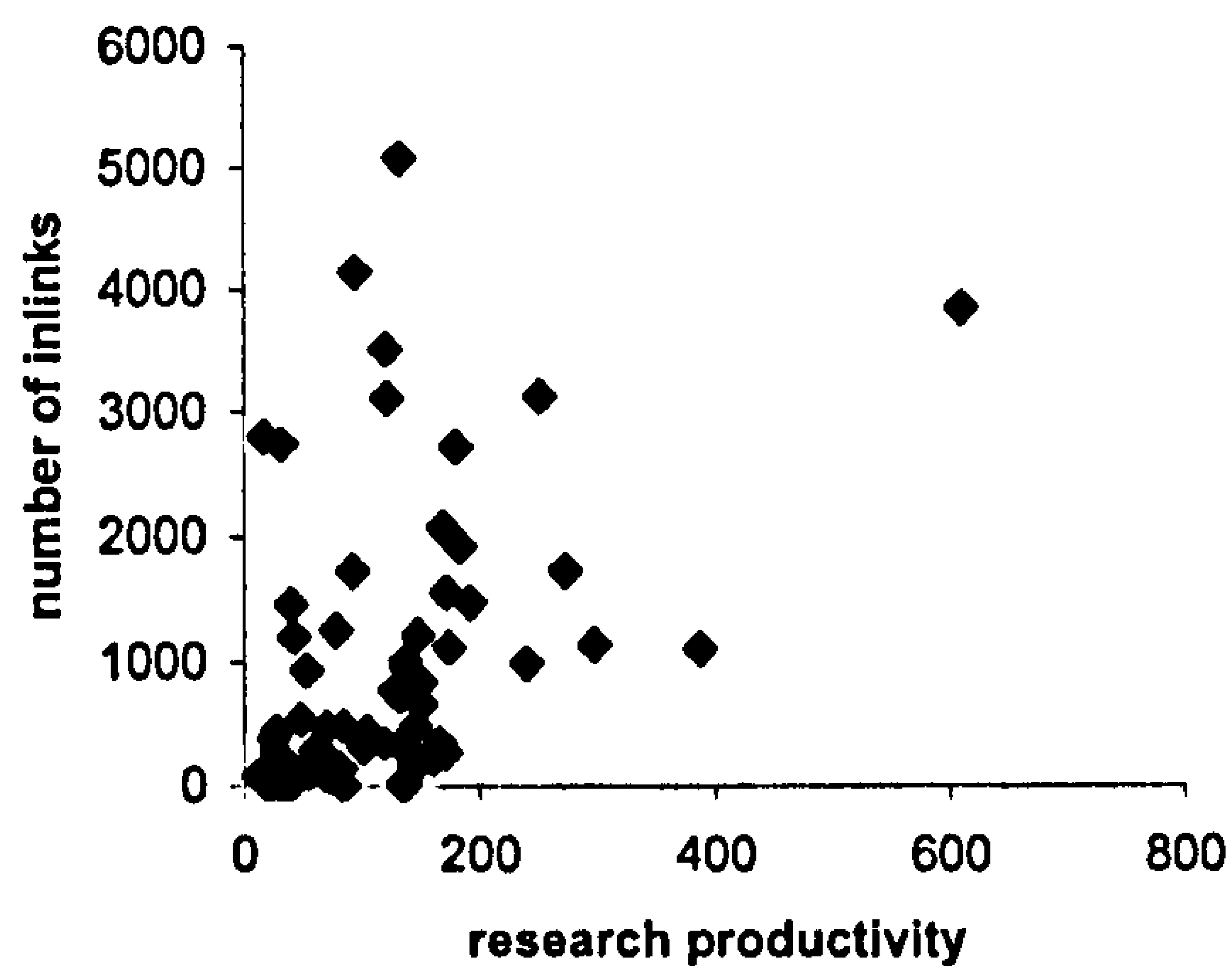


Figure 3.7 inlinks (from SocSciBot) against research productivities for the 79 UK computer science departments

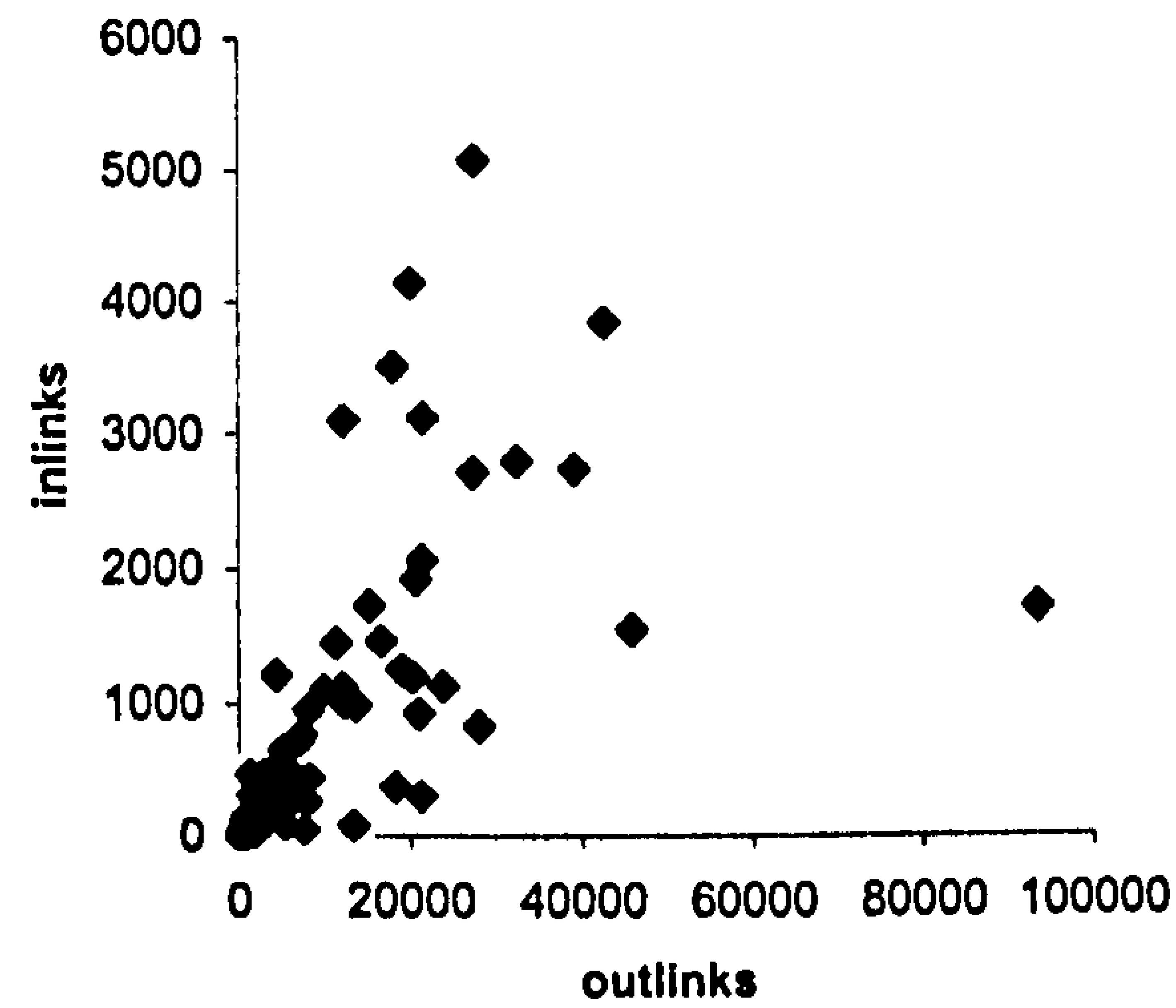


Figure 3.8 inlinks against outlinks (from SocSciBot) for the 79 UK computer science departments

3.1.3 Discussion

Significant correlations have been found between inlinks and research productivities, and between the WIFs (staff) and RAE averages for the 79 UK computer science departments, both from AltaVista and SocSciBot data. Although the results are not as good as those at university level, it is the first time statistically significant results have been found at the departmental level, and the positive results open the door to conduct further link analysis on departments. Comparing different WIF versions, the number of academic staff members is confirmed to be better than the number of web pages for the WIF denominator. The results from AltaVista are even better than those from the SocSciBot. This may be because that AltaVista has the ability to remember old domain names, although sometimes dead pages can also be included. This gives us some confidence that AltaVista is still a good choice in webometric research.

Two reasons may be causes of the positive results derived in this study. Firstly, the computer science departments make relatively more use of the Web than other disciplines, including library information science. Secondly, there is a more authoritative research assessment mechanism existing in the UK, the RAE rating, which decides where millions of pounds of research funding should go.

Compared with whole universities, the domain or directory names of computer science departments are inherently more complicated. A lot of computer science departments have several different domain names. In order to get a fairer link count to departments, old domain names are included in this study. If only the current domain names are used to obtain the number of links, some (especially departments that had recently changed their domain names) could hardly get any inlinks. For example, the computer science department in University of Bath was established from the department of mathematical science in August 2001. If only the new domain name '.cs.bath.ac.uk' were used to count the link pages from the .ac.uk domain, only 6 pages would be found, because many links still point to the old domain 'maths.bath.ac.uk'. When the old domain name was added to the query, more than two hundreds link pages were returned by AltaVista. Nevertheless, the academic staff members from the mathematics department are added when calculating the WIF (staff) for this department.

3.2 How Important Is It for Academic Web Sites in Western European Countries to Be in English in order to Attract Hyperlinks?

English has long been the major language at formal scholarly communication in the world. Journals indexed by the ISI are dominated by English, and citations can be lost to non-English journals (Moed, 2002). The WIF was first raised by Rodríguez Gairín (1997) in a Spanish documentation journal. Perhaps because it was not published in English, little attention had been paid to it until Ingwersen's (1998) paper 'The Calculation of Web Impact Factors' was published in an English journal: the Journal of Documentation. The analogous aim of this study is to find out whether English is important on the Web for attracting hyperlinks. Specifically, this experiment was designed to assess whether academic web sites in Western Europe attract more links if they are in English.

3.2.1 Method Adopted

AltaVista was chosen to collect data in this study, as it has a language facility for its advanced query. Part of the data is from a previous project by Thelwall et al. (2003b), which was collected in July 2002. In that study, the academic interlinking among sixteen Western European countries along linguistic lines was investigated. The focus in this study is on the English language impact regarding academic web interconnection, so only web pages and link pages in all languages and English were selected. Then the English proportions for academic web sites in each country were calculated through dividing the number of English web pages by the number of web pages in all languages.

$$\frac{\text{the number of web pages in English}}{\text{the number of web pages in all languages}}$$

The number of inlinks from other countries' academic web sites in all-languages and the number of web pages in all language for each country were used to calculate WIFs and link propensities.

WIF:

$$\frac{\text{the number of inlinks in all-languages}}{\text{the number of web pages in all-languages}}$$

Link propensity:

$$\frac{\text{the number of inlinks from another country's academic web sites}}{\text{the product of the web pages from the source and target countries' universities}}$$

The syntax: domain:.xx (xx stands for a country domain), was used to count numbers of web pages both in English and all languages for the sixteen Western European countries. The proportions of English web pages for each country are then calculated. The proportions are compared with those from academic web areas for each country, to put the study in context.

The Spearman rank order correlation coefficient was used to calculate the correlations between WIFs or link propensities and English page proportions of the sixteen Western Europe countries' academic web sites. Spearman rather than Pearson is chosen in this study, as the link data does not follow a normal distribution.

3.2.2 Results

Table 3.4 illustrates the English page proportions in both academic areas and whole country domains. English page proportions are much higher in the academic area than for the whole country domain, with one exception: Ireland. English is more heavily used in academia than other areas. Inlinks are derived from the data collected by AltaVista mentioned above. The Gross Domestic Products (GDPs) and the population for each country are from the CIA World Factbook 2000 (CIA, 2001).

Table 3.4 English proportions for the 16 Western European countries both in academic area and whole country domain

Country	Country Domain	Academic English proportion	Country English proportion	Inlinks in all languages	GDP (1000$)	Population (1000)
France	fr	0.43	0.18	29113	1382389000	59330
Italy	it	0.37	0.18	26517	1233324800	57632
Germany	de	0.51	0.21	97900	1879491900	82797
Spain	es	0.26	0.15	24260	691948100	39997
Greece	gr	0.14	0.20	6189	147367800	10602
Norway	no	0.39	0.17	17817	112473100	4481
Netherland	nl	0.37	0.24	48337	367105200	15892
Portugal	pt	0.53	0.32	8154	153734400	10048
Switzerland	ch	0.44	0.37	22416	196800200	7262
Belgium	be	0.45	0.32	20837	244759900	10241
Denmark	dk	0.68	0.28	12922	126996800	5336
Austria	at	0.53	0.20	29419	190265400	8131
Sweden	se	0.58	0.35	28988	183671100	8873
U.K.	uk	0.99	0.99	176934	1297339800	59511
Ireland	ie	0.97	0.98	14036	77079100	3797
Finland	fi	0.40	0.36	16267	108507000	5167

As shown in table 3.5, and in contradiction to previous research, the WIFs do not correlate significantly with English proportions. The correlation coefficient value is only 0.492, and not significant even at the 5% level. Figure 3.9 illustrates the relationship between the WIFs and English proportions. The reason may well be the unreliable choice of denominators (the number of web pages) in the WIF calculations. If the number of academic staff members of the universities were used as the denominators, positive results may have been obtained.

The link propensities correlate significantly with English proportions at the 1% level. However, the correlation coefficient value is low at 0.365, as shown in table 3.5. For the 16 countries, the link propensities were calculated for 240 pairs. Although the value is lower, the greater number of data points allows it to be more significant. Figure 3.10 shows the relationship between the link propensities and English proportions for the sixteen Western European countries. Once the impact of the sizes of both the source and target are removed, the link propensities correlate significantly with English proportions. That means English is apparently a contributing reason to attract links to web sites in Western Europe, although this may not be the only factor.

The Spearman correlation between inlinks and outlinks is high at 0.888, as shown in table 3.5, which is significant at the 1% level. Once both inlinks and outlinks are divided by the numbers of all language web pages, the correlation coefficient value between WIFs and WUFs is 0.759, which is still significant at the 1% level. This suggests that within the sixteen Western European countries, the more outlinks one country makes to the rest of the other countries, the more inlinks it receives.

It is also very impressive to see that the number of inlinks in all languages to a country correlates significantly with the number of English pages in that country (the Spearman correlation coefficient value is 0.868, and significant at the 1% level). This perhaps could lead to a conclusion that the more English pages in a country, the more inlinks it tends to attract from other countries. The high correlation coefficient value between the English pages and all language pages, can be interpreted as the higher ability to create web pages in a country, the more often English will be used. This also illustrates that English is the dominant language on the Web. The correlations between inlinks and all language pages, and between inlinks and English pages are all significant at the 1% level. However, the significant value may be caused by the underlying size effect. The number of all language pages in a country can be regarded as the size of a country's academic web sites. As mentioned above, once inlinks and English pages are divided by all web pages, there is not a significant correlation between WIFs and English proportions.

Population can also represent a nation's size, thus the significant results between inlinks and population, and between inlinks and GDP may also be caused by the underlying size effect. As shown table 3.5, when both inlinks and GDP are divided by the population in each country, there is not a significant correlation between inlink and GDP per person.

Table 3.5 Spearman correlation coefficient values for the sixteen Western European countries (* = significant at 5% level, ** = significant at 1% level)

Type of Correlation	Coefficient value	n
WIFs against English page proportions	0.494	16
Link propensities against English page proportions	0.365**	240
Inlinks against English pages	0.868**	16
English pages against all language pages	0.791**	16
Inlinks against all language pages	0.703**	16
Inlinks against population	0.621*	16
Inlinks against GDP	0.776**	16
Inlinks per person against GDP per person	0.238	16
Inlinks against outlinks in all languages	0.888**	16
WIFs against WUFs (all language pages as denominators)	0.759**	16

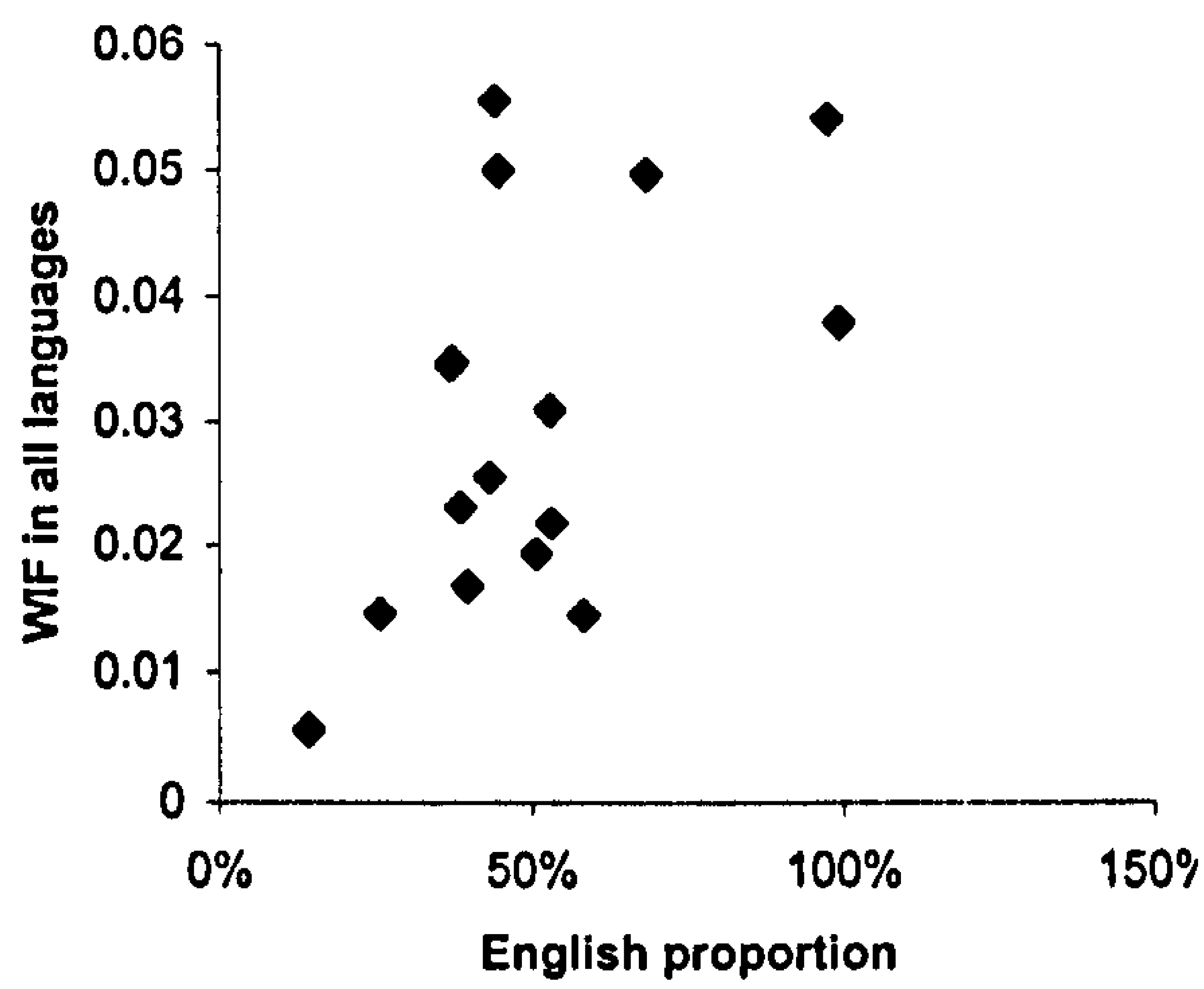


Figure 3.9 WIFs against English proportions for the sixteen Western European countries

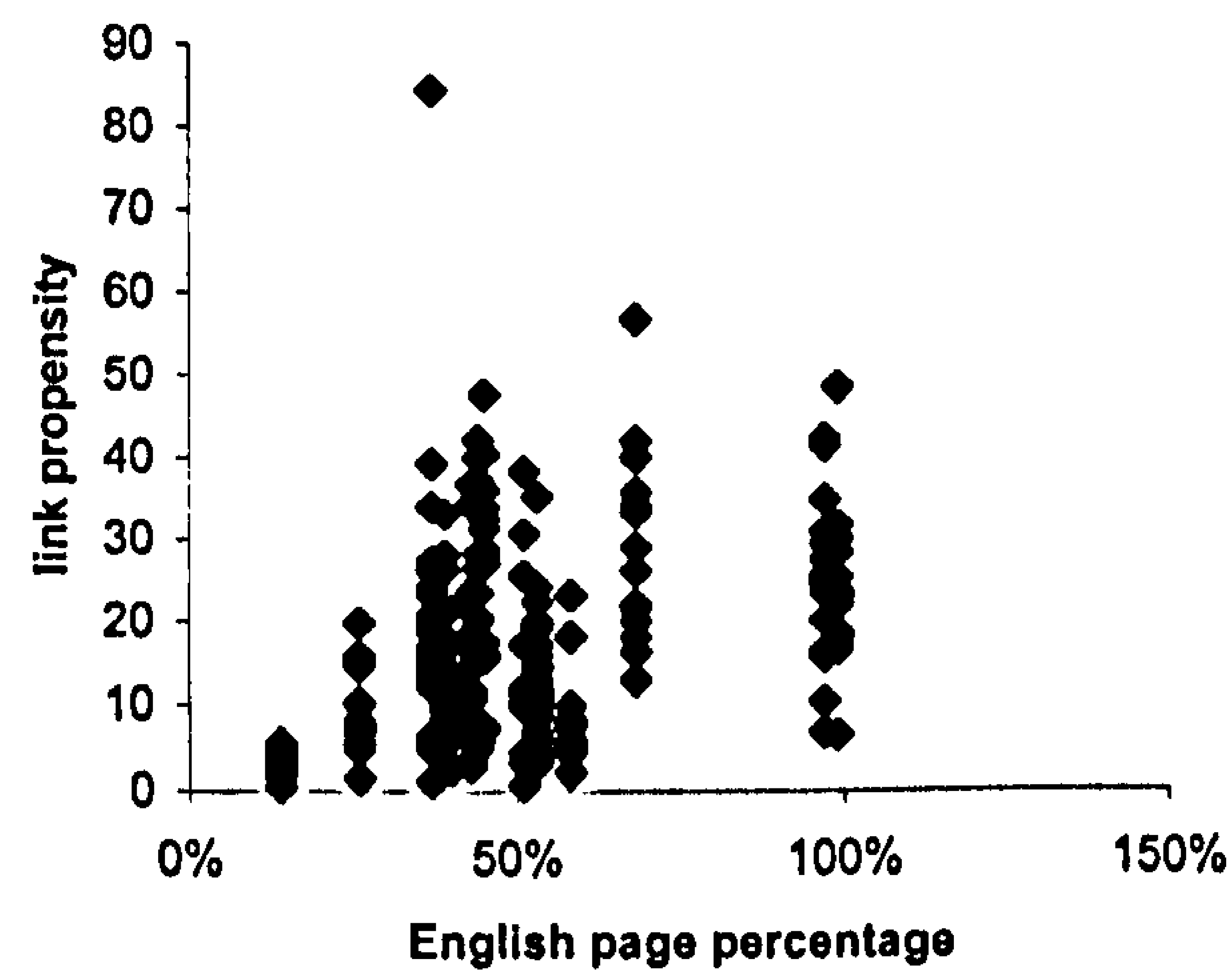


Figure 3.10 Link propensities against English proportions for the sixteen Western European countries

3.2.3 Discussion

The significant correlations identified between link propensities and English proportions give some evidence that English may be a factor for attracting hyperlinks, at least in Western European countries. Universities who do not want their ideas to be ignored on the Web should consider creating web pages in English as far as possible. The correlation coefficient value between WIFs and English proportions is nearly significant at the 5% level. This may be caused by the unreliability of the web page counts used in the WIF denominators.

There are some significant correlations between inlinks and English pages, all language pages, GDPs, populations. Nevertheless, all language pages and population represent to some extent a nation's size from different perspectives. Once the size effects are removed, hardly any significant results are found between WIFs and English proportions, and between inlinks and GDP per person. This may be a problem of the small sample size, however, only sixteen countries.

The low correlation coefficient value between link propensity and English proportion may suggest that although English is a reason to attract hyperlinks between those countries, it may not be the only reason. Other conditions such as technology levels, culture, history, politics and economic status may all play a role in this complex arena.

3.3 Summary

The significant correlation coefficient values between links and research measures for the 79 UK computer science departments, gave us some confidence to further investigate departmental interlinks. Some techniques applied for the computer science departments can be adopted in the main study to:

1. Collecting departments' domains or directories;
2. Collecting link data through both AltaVista and SocSciBot;
- Counting links from different domains through AltaVista
- Extracting departmental link data from existing universities' link data from SocSciBot
3. Conducting correlation tests.
- Inlinks versus research productivities
- WIFs versus research averages

The number of inlinks, which are attracted from other international peers within the Western European countries' academic web sites, are influenced by their English proportions and economic status. As a result, the main study selects samples from English speaking and similar economically advanced countries, as discussed in section 4.2.1.

4. Research Design and Methodology

This chapter discusses the research design which describes the

- Research hypotheses;
- Research design;
- Research methods.

Most of the current academic link analysis is at university level and focuses on the same country. Relatively less is known about departmental link analysis especially those across country boundaries.

This study aims at identifying departmental link patterns along country and disciplinary lines. It is necessary to conduct an experiment, firstly to validate the departmental link data, and secondly to find out whether there are country or disciplinary differences in national or international peer interlinking, and web use in general. Since it is difficult to identify authors of web pages, as discussed in sections 2.2.2 and 2.8.2, neither interview nor questionnaire survey can be applied in this study. For this reason, a quantitative research method was chosen (Tashakkori & Teddlie, 1998; Black, 1999).

4.1 Research Hypotheses

A hypothesis is a statement of expected research results or outcomes. It helps to determine the direction a study will take (Black, 1999). The hypotheses for this study are: departmental link data is a valid information source to disclose informal scholarly communication between departments; and departmental link patterns differ along both country and disciplinary lines. The following motivate the hypotheses.

4.1.1 Departments Are Better Research Units than Universities

The counterpart of link analysis, citation analysis at a lower aggregation level may disclose more information, since at a higher aggregation level, the result tends to average different citation uses amongst various subunits (van Raan, 2000). In this context, departments are better research units than whole universities to disclose online information.

There are some existing departmental level link studies:

- Thirteen Scottish computer science departments' links were compared with their organizational profiles with significant correlation coefficient values (Chen et al., 1998), although the departmental sizes were not taken into consideration.
- The web uses of library and information science departments from the UK (Thomas & Willett, 2000) and U.S. (Chu et al., 2002) have been investigated, although without significant correlations between inlinks and research ratings.
- Significant correlations were discovered between link metrics and RAE ratings for the 79 UK computer science departments (Li et al., 2003b), as described in section 3.1, although no link patterns were identified.
- Significant correlations between link metrics and estimated citation counts for the U.S. chemistry and psychology departments, while not for the history departments (Tang & Thelwall, 2003a). The link patterns were discovered to

be various between very different departments (Tang & Thelwall, 2004), although the study was conducted in one single country.

In summary, to conduct departmental link analysis is not only necessary but also plausible.

4.1.2 Link Data Validation

Both correlation tests and link motivation analysis have been used to validate link data, as discussed in section 2.6 and 2.8 separately. Correlation tests help to identify whether a set of link data is related to an existing data source of known value, while link motivation analysis identifies the nature of links, which can help to interpret information derived from the link data.

- As reviewed in sections 2.5 and 2.6, significant correlations between research measures and WIFs have been found at both university and departmental level in various countries.

- As reviewed in section 2.8.2.1, existing link motivation analysis, which was conducted at university level, discovered that university interlinking could disclose informal scholarly communication amongst universities.

Based on the above findings, the hypothesis is:

- **H1**: Links to departments associate with research in terms of a) significant correlations with existing research measures and b) a majority of links having some association with research.

4.1.3 Link Patterns Differ along National and Disciplinary Lines

In previous citation analysis studies, various disciplines or countries were discovered to use citations differently (Garfield, 1999; Glänzel, 2001; Glänzel & Schubert, 2001). It is natural to hypothesise that the same may be true for link uses. Previous link analysis discovered:

- Link uses are different with regard to disciplinary variations:
  1. Departments from hard sciences interlink more than those from social sciences (Tang & Thelwall, 2004).
  2. Journal web sites from library and information science make more use of the Web than those from law (Vaughan & Thelwall, 2003).
  3. Physics web pages make more intra-subject links than those in maths and sociology (Harries et al., 2004).
  4. Science and engineering dominate university web presence amongst the UK universities (Thelwall & Price, 2003).

- Link uses are different with regard to country variations:
  1. Norway was found to attract more inlinks per web page than other Nordic countries by Ingwersen (1998).
  2. Universities from New Zealand are less visible on the Web compared with those from Australia and the UK (Smith & Thelwall, 2002).
  3. Among Asia-Pacific universities, Australian and New Zealand are found to interlink heavily, and Australia and Japan were found to be the centre of the web use in that area (Thelwall & Smith, 2002).

iv. Universities from richer countries tend to make more use of the Web amongst the fourteen Western European countries (Thelwall et al., 2002).
v. For two geographical regions, Asia and Europe, European countries, especially the UK, Germany, Holland and Belgium were found to dominate the results. Chinese universities' web sites attracted significant number of links, and seem to be becoming the centre of the region (Park & Thelwall, 2006, to appear).

The above findings leave some key gaps in what is known, which the two hypotheses below seek to fill:

- **H2**: Departmental link patterns differ along country lines;
- **H3**: Departmental link patterns differ along disciplinary lines, even for similar disciplines.

4.2 Research Design

In order to pictorially represent the research design in this study, the notion of a variable map, which is adopted from concept maps (Novak & Gowin, 1984), is applied. The three graphical objects involved are:

- A rectangle represents an independent variable;
- A rounded box represents a dependent variable;
- An ellipse represents an extraneous variable.

A variable is defined as a general class of objects, events, situations, characteristics and attributes that are of interest to a study (Balnaves & Caputi, 2001). An independent variable does not depend on others, while a dependent variable does. An extraneous variable is a competing independent variable, which can influence dependent variables, but is not of interest to a study. Thus, extraneous variables should be controlled as much as possible. This means that either the value of an extraneous variable must have no effect to the result of a study, or is included in the design as an independent variable.

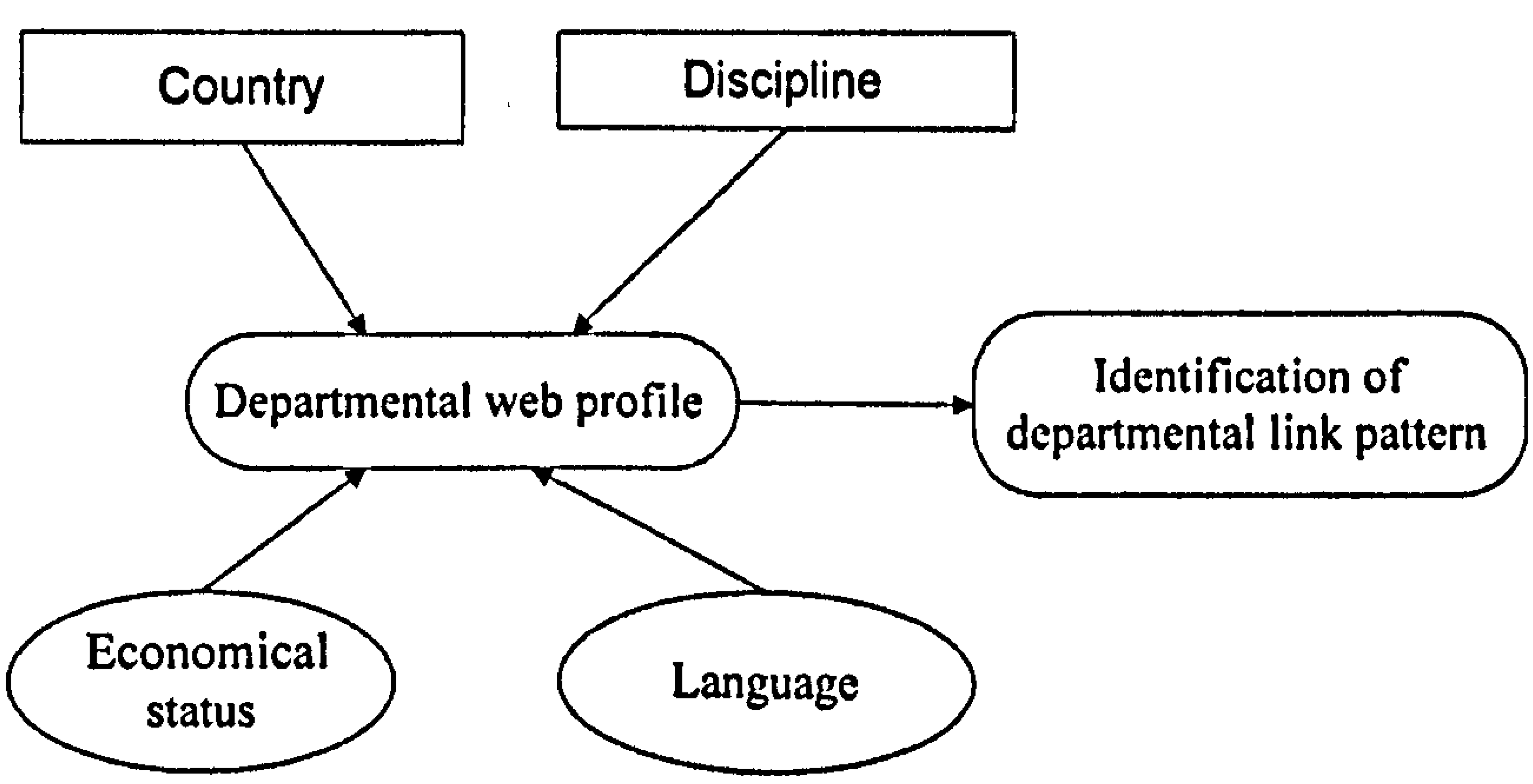


Figure 4.1 An example about variables

Figure 4.1 gives a simple example about different variables based on the concept maps. 'Departmental web profile' is defined as how a department use the Web with regard to publishing web pages, receiving and creating links. It is affected by independent variables: 'country' and 'discipline', and extraneous variables:

'economic status' and 'language'. 'Country' and 'discipline' are included in the research design by selecting various populations as described in section 4.2.1, while 'economic status' and 'language' were controlled through selecting departments from English-speaking countries with similar economic status. Research design may have limitations since not every extraneous variable can be controlled.

With regard to the two main objectives as described in section 1.3, sections 4.2.2 and 4.2.3 design two structures separately to:

- Validate departmental link data
- Identify departmental link patterns

4.2.1 Populations Selected In This Study

A population is defined as a group which share a set of common traits (e.g. all physics departments in the UK), while a sample is defined as a unit of analysis that is a subset of a population (Black, 1999; Balnaves & Caputi, 2001).

The populations of this study are nine sets of departments from physics, chemistry and biology disciplines in Australia, Canada and the UK, as listed in table 4.1. The populations can represent the two-dimensional independent variables 'country' and 'discipline' in this study as shown in figure 4.2 below. It is also possible to compare link patterns along country and disciplinary lines through the selected populations.

- Within each country, link patterns can be compared amongst different disciplines.
- Within each discipline, link patterns can be compared amongst different countries.

Table 4.1 Departments selected in this project for Australia, Canada and the UK

Australia	Canada	UK
Physics	Physics	Physics
Chemistry	Chemistry	Chemistry
Biology	Biology	Biology

A decision is made to use whole populations rather than samples in this study. This is because firstly most of the populations contain less than 50 departments which are manageable; secondly whole populations rather than samples ensure the external validity, since if a sample cannot represent the whole population where it is extracted from, the research design lacks external validation (Black, 1999).

Reasons for selecting departments from the three disciplines and countries:

- Physics, chemistry and biology are chosen because they are similar hard science disciplines. The following two points subsequently explain why similar departments, and why departments from hard science are of interest in this study:

i. Very different disciplines have been found to use the Web variously, with hard science making more use of the Web than those from social science.

This study focuses on similar departments to see whether there are still some disciplinary differences on web uses.

ii. Departments from social science, such as the U.S. history departments, may use the Web too little to conduct meaningful link analysis (Tang & Thelwall, 2004). As a result of this, a decision is made to choose departments from hard science to ensure enough number of links for investigation.

- Australia, Canada and UK are chosen because they are all economically advanced and English speaking countries. This can control the effects from the two extraneous variables: 'economic status' and 'language', as shown in figure 4.1.

By selecting the nine sets of departments from similar countries and disciplines in this study, it is expected that the outcomes can be extrapolated to a broader situation. If the different link patterns identified are consistent along country and disciplinary lines in this study, those from dissimilar countries and disciplines can be different to an even greater extent.

4.2.2 Validation of Departmental Link Data

A validation of departmental link data is to find out whether the data set is valid to disclose informal scholarly communication between departments. This is very important before any conclusion can be made from the departmental link analysis. As mentioned in section 2.1, both correlation tests and link motivation analysis have been used for the validation purposes. Figure 4.2 illustrates the design structure for this.

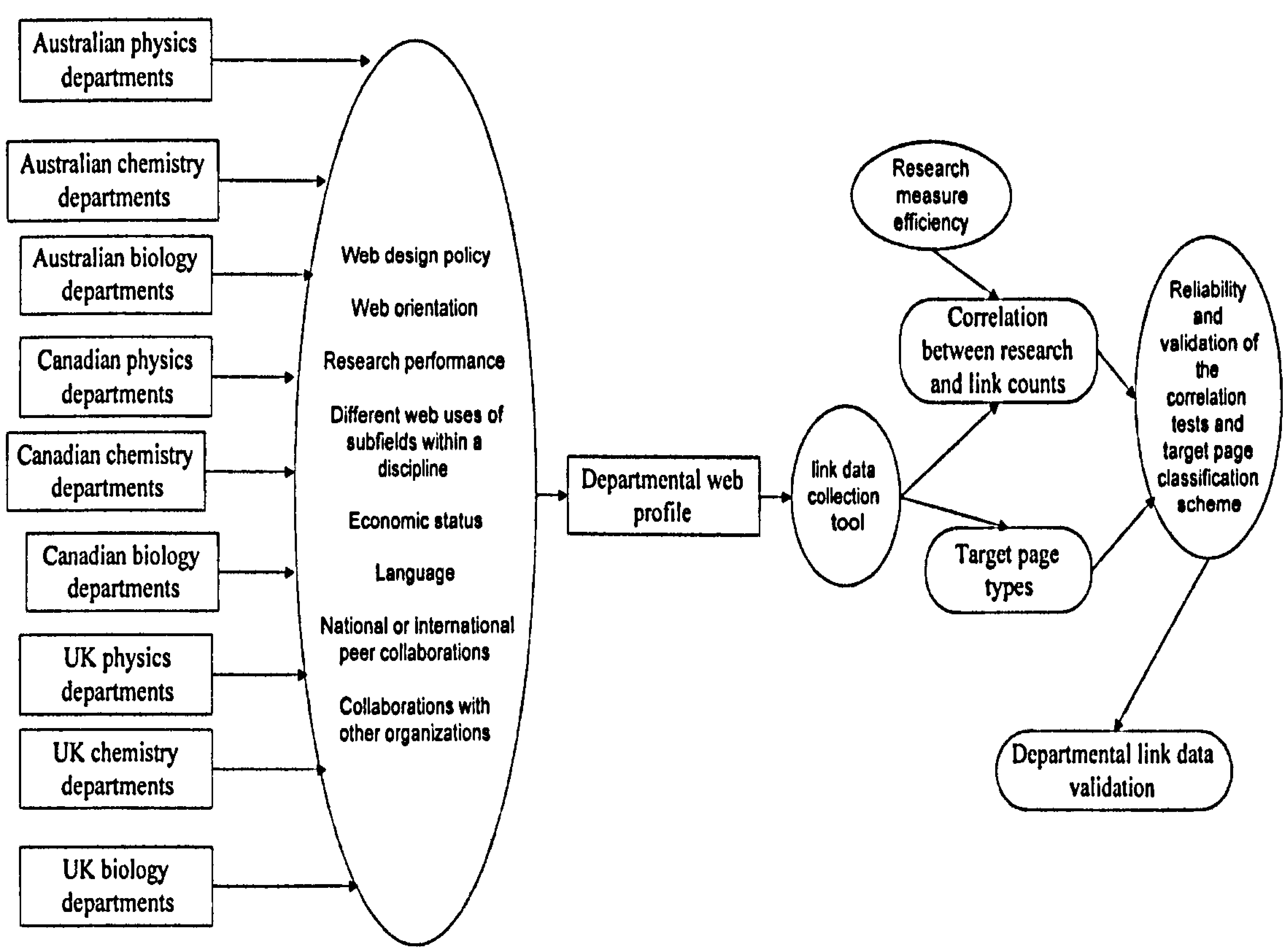


Figure 4.2 Link data validation

As shown in figure 4.2, the dependent variable is 'Departmental link data validation', while the operational definition of the dependant variable is 'correlation between research and link counts' and 'target page types'. In other words, the departmental link validation is through the correlation tests between research measures and link counts, and a target page type classification scheme. If the correlations are significant and the target page types are academic related, then the departmental link data are valid to disclose useful information. The design is the most comprehensive one of this type, since previous studies either focus on correlation tests as discussed in section 2.6, or focus on link motivation analysis as discussed in section 2.8.2.

The nine sets of populations on the left are used to represent the independent variables 'country' and 'discipline'. As the independent variable 'departmental web profile' for each set of departments is the same, only one is drawn in the diagram as a representative for all. Extraneous variables, which may affect how each set of departments employ the Web or 'departmental web profile', are listed below:

- 'Web design policy'. This has a direct effect on how well a department's web site can be crawled, since search engines or a self-designed web crawler tend to ignore links that are embedded within programs such as JavaScript or Shockwave.
- 'Web orientation'. Some departments tend to publish more on the Web than others.
- 'Research performance'. The better research performance of a department, the more web pages created and more inlinks attracted (Thelwall & Harries, 2004).
- 'Different web profile of subfields within a disciple'. A department's web profile may still be an average of various web uses from different subfields within it.
- 'Economic status'. Departments that have better economic status normally have more resources and make more use of the Web.
- 'Language'. Web pages in English tend to attract more inlinks as discussed in section 3.2.
- 'National or international peer collaborations'. Collaborations with national or international peers may result in more national or international peer link interactions.
- 'Collaborations with other organizations'. For example, departments collaborate more with industry may create or attract more links to or from com domain.

The extraneous variables for 'correlation between research and link counts' are:

- 'Research measure efficiency'. As discussed in section 2.7, various research measures have different limitations in evaluating departments' research performance accurately. In addition to citation counts, which are regarded as the most relevant research measures in the correlation tests, link counts are to be compared with other available research measures: NSERC research grants and RAE ratings are used for Canadian and UK departments respectively. Should the correlations between link counts and various research measures were all significant, one could have more confidence that link counts related

with research in different aspects. This serves as the control of 'research measure efficiency' extraneous variable.
- 'Link data collection tool'. The collection tools used, as discussed in section 2.3 either search engine or a self-designed web crawler, may bias the link counts. This is difficult to control, since search engines are good at collecting link data over large web areas while a self-designed crawler is good at covering individual web site extensively. Sometimes only search engines can be used, while at other occasions only a self-designed crawler is suitable. However, same types of link data are to be collected by same tools across all sets of departments. Hopefully this will average out the effects caused by the collection tools.

In addition to this, 'research performance' is not the only variable which may affect the 'departmental web profile', as mentioned above. Thus the correlations between research measures and link counts can indicate the relationship between 'research performance' and link counts, but cannot catch the relationships between other extraneous variables and link counts.

The extraneous variables for 'target page types' are:
- 'Link data collection tool'. The 'link data collection tool' used may bias the target pages identified. Hopefully that only SocSciBot is used to identify the national peer interlinking target pages for each set of departments may average the bias out.
- 'Target page disappearance'. By the time the target pages are visited, they may already disappear. This is to be controlled by defining the extraneous variable as one category in the target page classification scheme.
- 'Link context'. Identifying only target page types is not enough to disclose the nature of links. As discussed in section 2.8.2, the most comprehensive approach was devised by Bar-Ilan, which takes into account source, target page types together with link context analysis. Visiting source pages when necessary are to be used to control this extraneous variable.
- 'Other target pages'. Only national peer interlinking target pages are to be studied. As a result, other types of target pages are not known, such as web pages targeted by various TLDs, since it is not plausible to do all in this study.

The fact that not every extraneous variable can be controlled efficiently indicates that the design only partially validate the departmental link data.

4.2.3 Identification of Departmental Link Patterns along Country and Disciplinary Lines

Figure 4.3 illustrates that the departmental link patterns are identified along country and disciplinary lines, with regard to four different aspects. For simplicity, figure 4.3 does not show the nine sets of departments, since they are the same as those in figure 4.2.

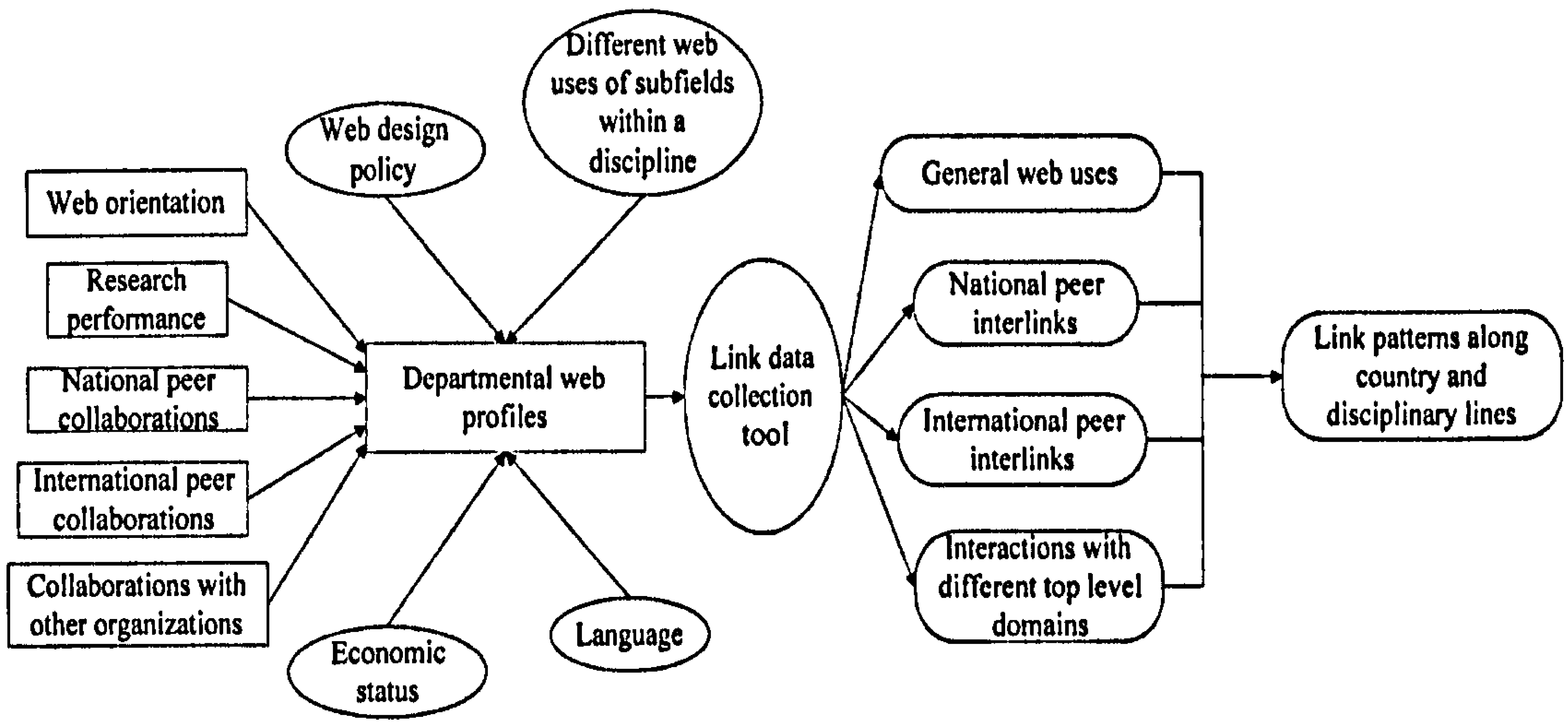


Figure 4.3 Link pattern identification

The dependent variable is 'link patterns along country and disciplinary lines'. The operational definitions of the dependant variable are:

- 'General web uses'. This is to measure a set of departments' web capacity, with regard to their abilities of publishing, making and attracting links on the Web. This is designed to catch the effects from 'web orientation' and 'research performance' variables. Departments that have higher research performances may tend to make more use of the Web, publishing more web pages and creating or attracting more links. The four values below are designed to indicate departements' general web uses:
  i. The number of web pages each department publishes.
  ii. The number of external outlinks made by each department.
  iii. The number of inlinks each department attracting from the rest of the Web.
  iv. The proportion of national inlinks for a set of departments, with regard to those from the whole Web.

- 'National peer interlinks'. This is to show how well a set of departments are interconnected, and also the ability of a set of departments to attract national peer inlinks, which are from the same type of departments in the same country. This has the potential to disclose the underlying 'national peer collaborations', where departments collaborate more may interlink more than those do not. The following two values are designed to show this aspect:
  i. Interconnection rate of each set of departments.
  ii. The number of national peer inlinks each department receives.

- 'International peer interlinks'. This is to show a set of departments' ability to attract international peer inlinks, which are from the same type of departments in two other countries. This is designed to catch the underlying effects from 'international peer collaborations'. Two values are designed as below:
  i. Link Propensities for each set of departments.
  ii. Adapted mean international peer inlinks.

- 'Interactions with different top level domains'. This has the potential to illustrate how a set of departments interacts with other organisations on the Web ('collaboration with other organiztions'). This is an analogy as patent citation analysis between patents to scientific literature to disclose science and technology interactions, as reviewed in section 2.9.2. For this purpose, two proportion values are designed as follows:
  i. Proportions of external inlinks from different top level domains with regard to those from the whole Web.
  ii. Proportions of external outlinks to different top-level domains with regard to all external outlinks made.

Section 5.5 explains in detail how each value is computed.

The extraneous variables 'economic status' and 'language' are controlled through choosing departments from similar economically advanced and English speaking countries as discussed above. However, extraneous variables 'web design policy' and 'link data collection tool' may bias the link data collected, since web sites, which use advanced technology such as Frameset, JavaScript and Shockwave, may not be covered fully by either search engines or a self-designed web crawler. 'Different web uses of subfields within a discipline' indicate that the departmental link pattern identified is still an average value although it is more specific than that of whole universities.

In summary, through the four aspects, a whole picture of departmental web uses can be captured. However, the uncontrollable extraneous variables 'web design policy' and 'link data collection tool' may distort the results.

4.3 Research Methods

4.3.1 Research Procedure

Given the research hypothesis and the two research designs, the general research procedure is defined as follows:

1. Collecting link data through both AltaVista and SocSciBot. The literature reviews that search engines are good at covering large web areas, while a self-designed web crawler is good at covering small web areas thoroughly. As a result of this,

- SocSciBot is used to collect:
  i. Web pages each department publishes.
  ii. National peer inlinks
  iii. External outlinks made by each department
  iv. External outlinks to different top level domains
- AltaVista is used to collect:
  i. External inlinks from different top level domains and the whole Web.
  ii. International peer inlinks from two other countries.

2. Collecting some relevant non-web data for the departments.
   i. Existing research measures. As discussed in section 2.7.2.2, the RAE scores can be used to represent UK departments' research measures. The NSERC grants can be used to represent Canadian departments' research measures, as discussed in section 2.7.2.3.
   ii. Number of academic staff members. This is used to represent the size of a relevant department, as discussed in section 2.5.3. It is used to remove size effect from both link and research measures.
   iii. Citation counts. As discussed in section 2.7.2.1, citation counts are the most relevant data set for conducting the correlation tests for link data.

3. Conducting correlation tests.
   i. Correlation tests are conducted for each set of departments between citations and link counts.
   ii. Correlation tests are conducted between RAE ratings and link counts for the UK departments.
   iii. Correlation tests are conducted between research grants and link counts for Canadian departments.

4. Classifying target page types.

5. Computing different values with regard to the four aspects.

6. Identifying link patterns along country and disciplinary lines with regard to the four aspects.

The experimental procedure details, which are used to conduct the research step by step, are described in the next chapter.

4.3.2 Statistical Methods Used

The frequency of inlinks form a highly skewed distribution (Thelwall & Wilkinson, 2003a), for which Spearman is more appropriate than Pearson in the correlation tests. Nevertheless, both Pearson and Spearman formulae were used to calculate the correlation coefficient values in this study, to give a full picture of the relationship between the link based metrics and research measures. One can spot the apparent outliers more easily than just using Spearman. The statistical software SPSS is used to calculate both the correlation coefficient and median values in this research. The software Medcalc (2003) is used to calculate whether a difference between two sets of correlation coefficient values is significant.

4.4 Summary

A quantitative research method is chosen for this study. The research hypothesis is: Departmental link data is a valid information source to disclose informal scholarly communications, and link patterns are different along country and disciplinary lines. In order to test the hypothesis, nine sets of departments are chosen from physics, chemistry and biology in Australia, Canada and the UK.

The research design shows how to:

- Validate departmental link data
- Identify departmental link patterns along country and disciplinary lines

The validation of departmental link data is through correlation tests and target page type classification, while the identification of departmental link patterns is through analysing departmental web profiles from four aspects. The four aspects being:

- General web use
- National peer interlinks
- International peer interlinks
- Interactions with different top level domains

The general research procedure is defined as follows:

1. Collecting link data through both AltaVista and SocSciBot
2. Collecting some relevant out web data
3. Conducting the correlation tests
4. Classifying target page types
5. Computing different values with regard to the four aspects
6. Identifying link patterns along country and disciplinary lines

There are some limitations for the two research designs in this chapter. 'Research performance' is only part of the reasons for a department to attract inlinks. Thus the correlation tests between research measures and link counts can only partially indicate the relationship between research and link counts, but fail to identify the relationships with other extraneous variables. The identification of target page types also has limitations, since it is not the most comprehensive one and also the collection tool used may bias the target pages identified. 'Web design policy' and 'link data collection tool' may both affect the departmental link data collected. As a result the link patterns identified may be distorted by these extraneous variables. In summary, the validation design can only partially validate the link data; while when making conclusions on the link patterns identified, one should always bear the extraneous variables in mind.

5. Experimental Procedure

Based on the general research procedure introduced in section 4.3.1, this chapter explains how the experiment, which is used to achieve the two main objectives, is carried out in this study.

Particularly, this chapter describes how to:

1. Collect relevant link data through AltaVista and SocSciBot
2. Collect relevant non-web data, these include:
   i. Research measures
   ii. Number of academic staff members
   iii. Citation counts
3. Conduct correlation tests
4. Classify target page types
5. Compute different values with regard to the four aspects

5.1 Link Data Collection

The link data collection includes the number of web pages, inlinks and outlinks for each department. As discussed in section 4.3.1, SocSciBot and its processing software are used to collect

- Number of web pages each department publishes;
- Number of national peer inlinks each department receives (with four ADMs);
- Number of external outlinks made by each department;
- Number of external outlinks to different top level domains from each set of department,

while AltaVista is used to collect

- Number of external inlinks from different top level domains and the whole Web;
- Number of international peer inlinks for each set of departments from their peers in two other countries;
- Number of web pages each department publishes.

Before the link data can be collected, the departments' domain or directory names have to be identified. The three subsections subsequently explain how to

- Collect departments' domain or directory names;
- Collect relevant link data through SocSciBot;
- Collect relevant link data through AltaVista.

5.1.1 Collecting Departmental Domain or Directory Names

For the three countries, each university has one or more proper domain names. For example, universities in the UK have domains like: www.xxx.ac.uk. The universities are differentiated by the third level domain xxx. The domain names for universities in Australia appear as: www.xxx.edu.au. These universities are differentiated by the third level domain xxx. The domain names for universities in Canada have the form: www.xxx.ca. There is no special academic domain for Canada.

The departmental domains are more complicated. Some departments have proper domains which are derivable from their parent universities. For example, www.phys.ualberta.ca is the domain name for the physics department from the University of Alberta in Canada. Some departments' urls are just directories in their parent universities' domains. For example, www.nottingham.ac.uk/physics is the url for the physics department from the University of Nottingham in the UK. Some have their own and others share urls. For example, www.spme.monash.edu.au is the domain name for the school of physics and materials engineering from the University of Monash in Australia, and there is not a separate domain name for the physics department. The domain or directory names for departments are identified from the Web manually.

5.1.1.1 Identifying the domain or directory name lists for the departments in Australia

The physics, chemistry and biology departments found in the citation and publication data supplied by the Research Evaluation and Policy Project (REPP), Research School of Social Sciences, Australian National University, are included in this project. The domain or directory names for those departments are then identified manually through the search engine Google. If any research groups' domain or directory names are found out to be different from their own departments', they are added to the relevant domain or directory name list. The domain or directory name lists are also compared with several physics, chemistry and biology departments' link lists found using Google. More domain and directory names are identified using this extra source.

5.1.1.2 Identifying the domain and directory name lists for the departments in Canada

The physics, chemistry and biology departments, which have received research grants in 2003 from National Sciences and Engineering Research Council of Canada, are included in this project. Again, the domain and directory name lists for these departments are identified manually through Google and compared with several relevant departments' link lists.

5.1.1.3 Identifying the domain or directory name lists for the departments in the UK

The physics, chemistry and biology departments submitted in the RAE 2001 are included in this project. From the homepage of RAE 2001 http://www.hero.ac.uk/rae/, following the menu 'Submissions' and then 'View by Unit of Assessment', physics, chemistry and biology departments are identified from '19: Physics', '18: Chemistry' and '14: Biological Science', respectively. Research groups submitted are checked to make sure the right departments are identified. Just as in Australia, the domain or directory name lists for these departments are identified manually and then compared with several relevant departments' link lists.

5.1.2 Collecting Link Data through SocSciBot3

Whole university's link data is obtained first for the three countries. For Australia, forty-two Australian universities' domains were identified and crawled in June-July 2003 by the SocSciBot3. For Canada, ninety-seven universities' domains were identified and crawled in January 2004. However, only ninety universities' link data has been collected. The other seven have been removed because their sites could not be crawled. For the UK, one hundred and twenty five universities were crawled in June 2003 and the link structure database was downloaded from an online source (The Statistical Cybermetrics Research Group, 2003).

After all university link data is collected, the link data for physics, chemistry and biology departments in Australia, Canada and the UK was extracted by a program called 'subsite extractor' together with the relevant domain or directory name lists identified above. The approach is the same as that used in the previous study about the 79 computer science departments, as discussed in section 3.1. The program can extract the link structure which contains only the domains or directories of a department in the source pages from the relevant university. Tables 5.1 and 5.2 below gives an example on how the 'subsite extractor' works. Table 5.1 lists the university and departmental domain names of Birmingham University. In table 5.2, the left column gives a simple example of how a university's link structure is organized, which is crawled by SocSciBot. Each url, which starts from the beginning of each line, represents a web page within the web site in question, while those urls which follow a tab space above a web page url, are the links collected on the web page. The right column illustrates the departmental link structures for the physics, chemistry and biology departments from the top to bottom row. For physics department, only the web page contain its domain name 'ph.bham.ac.uk' is selected. The same is for the chemistry and biology departments. The web page '.bham.ac.uk/' is ignored by the 'subsite extractor', since it is not from any of the three departments' domains.

Table 5.1 University and departmental domain names for University of Birmingham

University domain name	bham.ac.uk
Physics department	ph.bham.ac.uk
Chemistry department	chem.bham.ac.uk
Biology department	biosciences.bham.ac.uk

Table 5.2 An example of extracting departmental link data from university link data

University link data	Extracted departmental link data
.webteam.bham.ac.uk/accessibility.htm .general.bham.ac.uk/contact.htm .bham.ac.uk/	.ph.bham.ac.uk/staff/ .general.bham.ac.uk/legal/privacy.htm .ph.bham.ac.uk/
.ph.bham.ac.uk/staff/ .general.bham.ac.uk/legal/privacy.htm .ph.bham.ac.uk/	.chem.bham.ac.uk/staff/ .chem.bham.ac.uk/staff
.chem.bham.ac.uk .chem.bham.ac.uk/staff .general.bham.ac.uk/legal/privacy.htm .biosciences.bham.ac.uk/research/	.general.bham.ac.uk/legal/privacy.htm .biosciences.bham.ac.uk/research/

By collecting departmental link data this way, existing university link data can be used to extract relevant departmental link data without repeat crawling. In this study, the UK departments are mainly extracted from the existing university link data. Another benefit is that by extracting link data from relevant university's link structure, more departmental web pages can be crawled than just crawling departmental web site separately. This is because those departmental web pages, which cannot be followed from a department's homepage, but can be following from its university homepage, are included in the university link data collection.

Should no link data be extracted, the department is crawled again separately using the SocSciBot3. If a department cannot be crawled, it is simply removed from the project. In total, 326 departments' link structures were derived, while altogether 32 departments have been removed from this study.

Once the departmental link data has been collected, it can be processed using the relevant software, as described below. The software belongs to the Statistical Cybermetrics Research Group in the University of Wolverhampton (2005).

- To analyse each department's link structure. The number of web pages and external outlinks for each department are calculated by a piece of software.
- To count links between departments in the same set with different ADMs. The existing software to count links with different ADMs works at the university level. The departments' link structures have to be first extracted in order that only the relevant target urls from the same set of departments are left. Then the links to a set of departments from their peers in the same country can be counted with different ADMs as if they are different universities. The link counts from the file ADM is the national peer inlinks.
- To analyse how a set of departments link to various TLDs. A piece of software is used to summarise which top level domains a set of departments links to and how many links target each top level domain.

5.1.3 Collecting Link Data from AltaVista

AltaVista's advanced search facility, which has been stopped completely (Pandia, 2004), are used to count links from large web areas. The domain and directory name lists collected above are used to collect the departmental link data. Below, some examples of syntax, which were used to count links using AltaVista's advanced query facility, are listed.

- To count links from the whole Web to a department:

LINK: phys.ualberta.ca AND NOT HOST:ualberta.ca

This syntax is used to count the links from the whole Web to the physics department of the University of Alberta in Canada. The syntax 'AND NOT HOST:ualberta.ca' is used to exclude links from the same university. 'LINK' and 'AND NOT HOST' are in uppercase to emphasize that they are keywords. However, it is not obliged to type it in uppercase in order to count the number of links in AltaVista.

- To count links from a domain to a department:

LINK: phys.ualberta.ca AND DOMAIN:com

or

LINK: phys.ualberta.ca AND DOMAIN: ca AND NOT HOST:ualberta.ca

The first syntax above is to count the links from com domain to the physics department from the University of Alberta in Canada. The second syntax is to count the number of links from ca domain to the same department. 'AND NOT HOST:ualberta.ca' is the syntax to make sure that no internal links from the same university are counted to this department.

- Counting links between two sets of departments from the same discipline in two different countries:

The number of links from one set of departments in one country to each department from another set of departments in the same discipline in another country is counted. The sum of the results is the number of links from the first country to the second country in that discipline. The syntax for AltaVista is:

HOST: phys.ualberta.ca AND LINK:bath.ac.uk/physics

This syntax is used to count the number of links from the physics department of the University of Alberta in Canada to the physics department of the University of Bath in the UK.

- To count the number of web pages within each department

HOST: phys.ualberta.ca
or
URL: bath.ac.uk/physics

The first syntax is used to count the number of web pages for those departments who have a proper domain name. The second form of syntax is used to count the number of web pages for departments with directory names.

Table 5.3 lists domains from which the links are counted to each country's departments by AltaVista. Recall that unlike the UK and Australia, Canada does not have its own commercial and academic domains.

Table 5.3 The domains that links are counted from

Country	Domains from
Australia	The whole Web, com, edu, au, org, net, com.au and edu.au
Canada	The whole Web, com, edu, ca, org and net
UK	The whole Web, com, edu, uk, org, net, ac.uk and co.uk

Unfortunately after the link data had been collected for this project in late March 2004, the advanced query facility of AltaVista was stopped completely.

5.2 Non-Web Data Collection

As discussed in section 2.7, research measures are needed in order to carry out the correlation tests for departmental link data validation. Number of full time academic staff members for each department is also needed to factor out the size effect from both link and research measures.

This section describes, how relevant research measures and number of academic staff members are collected. Particularly, this section explains how citations are counted for the Canadian and UK departments in a semi-automatic way.

5.2.1 Research Grants for Canadian Departments

As discussed in section 2.7.2.3, the 2003 research grants competition results conducted by the National Sciences and Engineering Research Council of Canada (NSERC, 2003) has been used to represent the research quantitative indicators for physics, chemistry and biology departments in Canada. The grants received by each department have been accumulated. Biochemistry departments are not included as according to Leydesdorff (2004) biochemistry neither belongs to biology nor chemistry. Chemistry and biochemistry departments, however, are included in chemistry departments, while biology and biochemistry departments are included in biology departments.

5.2.2 RAE Ratings for UK Departments

Many studies have demonstrated that there is a strong correlation between citation counts and RAE ratings (Oppenheim, 1995; Oppenheim, 1997; Smith & Eysenck, 2002; Norris & Oppenheim, 2003). The ratings for RAE 1992 are from 1 to 5, while the ratings for RAE 1996 and 2001 are: 1, 2, 3b, 3a, 4, 5, 5* (HEFCE, 2005). In order to conduct the correlation tests, the latter ratings were replaced by 1, 2, 3, 4, 5, 6, 7 respectively. In those studies, each department only has one rating value. In this study, under the Biological Science Unit of Assessment (UoA) of the RAE 2001, a lot of departments have more than one rating value. For example, the University of Cambridge has five different ratings for Biochemistry, Genetics, Plant Sciences, Zoology and Biotechnology separately. One average RAE value for each department is necessary to conduct the correlation tests. Another reason for an average RAE score is to give each department a fairer value. The proportions of active research staff members submitted are different across departments. Some may submit small proportions to aim higher RAE ratings for better reputation; the others may do the opposite to receive more research funding, as the amount of funding is determined by both RAE rating and research active staff members submitted (HERO, 2001c).

The technique, which was reviewed in section 2.7.2.2, was used to calculate average RAE scores for each department. In order to justify that approach, two types of correlation tests are conducted between the research productivities (numerator part of that formula) and relevant research incomes submitted in the RAE 2001, and between the average RAE scores and the research incomes per academic staff member. The research incomes are chosen for the correlation tests, since they can be used as research quantitative indicators of universities' research outputs, as discussed in section 2.7.1.2. Both Pearson and Spearman correlation coefficient values are significant at the 1% level for not only all institutions but also the departments in physics, chemistry and biology that submitted in the RAE 2001, as listed in Appendix E. The results from the formula somehow reserve the relative research quantities of the institutions and departments, even if the RAE scores by nature should not be mathematically manipulated in this way. They should be acceptable as proxies of research quantitative indicators in the correlation tests.

5.2.3 The Number of Academic Staff Members

The number of academic staff members for each UK departments are from the RAE submission, through accumulating 'Category A and A* Research Active Staff' and 'Non-selected Category A and A* Research Active Staff'(HERO, 2001d).

The number of fulltime academic staff members for Australian and Canadian departments are identified and counted manually from each department's web site, as no official information is available. A small number of departments do not provide their staff information on their web sites. These departments are ignored when the number of staff members is used to remove the size impact for their link and research measures. In order to count the staff members fairly, only fulltime professors, associate professors, readers, senior lecturers and lecturers have been included. This is because not every department lists their postdocs, research assistants, research associates or research fellows on their web sites. For Canadian departments in Quebec whose web sites are in French, help has been acquired from a French-speaking lecturer in Quebec to make sure the numbers are counted correctly.

5.2.4 Counting Citations from ISI's Web of Science

As discussed in section 2.7, citation counts are the most relevant data set for the correlation tests. Numbers of citations for Australian physics, chemistry and biology departments were supplied by the REPP, Research School of Social Sciences, Australia National University. The publication and citation counts come from the REPP database, which covers all Australian publications in ISI's three main indices: the Science Citation Index (SCI), Social Science Citation Index (SSCI) and Arts & Humanities Citation Index (A&HCI). For the SCI, the data has the reduced journal coverage similar to that on the CD-ROM version. The data included articles, notes and reviews from 1998 to 2002. The 2002 data is the latest with addresses that have been cleaned to the departmental level by the REPP.

However, the citation counts for Canadian and UK departments are not available. Section 2.7.1 proposes a semi-automatic technique to count the citations from ISI's Web of Science. This section explains the issue in detail.

5.2.4.1 Collecting Relevant ISI Records

The citations are counted for each department in a set. Relevant ISI records are collected for each of the three disciplines in Canada and UK respectively. The following three steps show how each set of departments' ISI records are collected.

1. Select relevant journal names for each of the three disciplines. Journal names, which are categorized as physics, chemistry and biology are collected separately based on 'Subject Category Selection' follow 'ISI Journal Citation Reports' from the Web of Science (ISI, 2005).
2. Search relevant ISI records for each set of departments. The following three steps are used to search relevant ISI records:
   i. Go to ISI Web of Science press 'Full Search' button, and set the time 'from 1998 to 2002';

ii. Press 'General Search' button, fill in 'Source Title' text box with the relevant journal names, 'Address' with country code. The country code for Canadian departments is simply 'canada'. The country code for the UK is: 'england or scotland or wales or north ireland', since this is how the country code is organized for the UK papers indexed by the ISI.
iii. Press the search button.

3. Save the searched ISI records in a plain text file. The following four steps are used to save the ISI records into a plain text file.
   i. In the 'General Search Results' window, press the 'Mark All' button;
   ii. Press 'Marked List' button, tick 'address' and 'times cited';
   iii. Press 'Format for Print' button';
   iv. Select and copy the records into a text file.

Each time only up to five hundred records can be collected this way. As a result, each set of departments' ISI records are collected repeatedly until all are included.

5.2.4.2 Counting Citations for Each Department

As discussed in section 2.7.1, the counting of citations for a department is by matching its postcodes with those in the address part of a ISI records. Each department's postcodes are identified through their departmental homepages manually. Details of the sets of departments' postcodes are listed in Appendix J.

A program is written to count citations automatically for each of the six sets of departments. For each department, the program checks from the first record until the end of the relevant ISI record file, if the address part of a record has the department's postcode, then the 'times cited' of that paper is accumulate to the amount of the citations that department receives, and the number of its publications is increased by one. For each matching paper, the citations are counted only once for the department, no matter how many authors in that paper are from the same department. In this way, the program computes citations and publications for each of the six sets of departments.

5.2.4.3 Validation of the Citation Counts

Several correlation tests are carried out between the citation counts and research measures for validation purposes. Table 5.4 shows that for Canada, research grants are used to compare with the citation counts, while for the UK, the RAE ratings are used. Citation counts, citations per staff member and citation per publication (CPP) are compared with relevant research measures.

Table 5.4 Citation counts correlation tests

Country	Citation	Research measure
Canada	Citation counts	Research grants
UK	Citation counts	Research productivity (RAE rating multiplied by the staff members submitted)
Canada	Citations per staff member	Research grants per staff member
UK	Citations per staff member	RAE average
Canada	CPP	Research grants per staff member
UK	CPP	RAE average

5.3 Correlation Tests

As discussed in section 2.6, the correlation tests between link and research measures are carried out to validate link data. Citations are regarded as the most relevant data set to compare with link data. Hence, correlation tests are carried out between citation and link counts at three different levels, as discussed in section 2.7. RAE ratings and research grants are used in the correlation tests for UK and Canadian departments respectively to further validate the departmental link data.

Table 5.5 illustrates the two sets of data that are used to carry out the correlation tests between the national peer inlinks with different ADMs from SocSciBot and their relevant research measures.

Table 5.6 shows the two sets of data that are used to calculate correlation coefficient values between the link measures from different web areas collected from AltaVista and those relevant research measures.

Table 5.5 National inter-departmental link correlation tests

Country	Link data	Research measure
Australia	Inlinks & outlinks with different ADMs	Citations
Canada	Inlinks & outlinks with different ADMs	Citations and research grants
UK	Inlinks & outlinks with different ADMs	Citations and research productivity (RAE rating multiplied by the staff members submitted)
Australia	WIFs & WUFs (staff members as denominators)	Citations per staff member
	WIF & WUFs (links with different ADMs as numerator, web pages, directories, domains and departments as denominators)	CPP
Canada	WIFs & WUFs (staff members as denominators)	Citations and research grants per staff member
	WIF & WUFs (links with different ADMs as numerator, web pages, directories, domains and departments as denominators)	CPP and research grants per staff member
UK	WIFs & WUFs (staff members as denominators)	Citation per staff member and RAE average
	WIF & WUFs (links with different ADMs as numerator, web pages, directories, domains and departments as denominators)	CPP and RAE average

Table 5.6 International link correlation tests

Country	Link sources	Research measure
Australia	Links from the whole Web, com, edu, au, org, net, com.au and edu.au domains	Number citations
Canada	Links from the whole Web, com, edu, ca, org and net domains	Citations and research grants
UK	Links from the whole Web, com, edu, uk, org, net, ac.uk and co.uk domains	Citations and Research productivity (RAE rating multiplied by the staff members submitted)
Australia	WIFs with staff member as denominators	Citations per staff member
	WIFs with web pages as denominators	CPP
Canada	WIFs with staff member as denominators	Citations and research grants per staff member
	WIFs with web pages as denominators	CPP and research grants per staff member
UK	WIFs with staff member as denominators	Citations per staff member and RAE average
	WIFs with web pages as denominators	CPP and RAE average

The research measures to be compared with the number of links are research productivities. For Australia, the numbers of citations are used to represent research productivities. For Canada, both citation counts and research grants received by each department are used to represent research productivities. For the UK departments, both citation counts and research productivities, which are calculated by multiplying their RAE ratings with the number of academic staff members submitted, are used to represent research productivities.

The research measures to be compared with the WIFs or WUFs are research averages. For citation averages, if the denominators for WIFs or WUFs are the numbers of full time academic staff member, the citation averages are citations per academic staff members, while when the denominators of WIFs or WUFs are the numbers of web pages, directories, domains and departments then Citation Per Publication (CPP) are used to represent the research averages. For RAE and research grants averages, no matter which denominators the WIFs or WUFs use, the research average has only one version.

5.4 Identification of Target Page Types

The departmental target pages derived from the link data crawled by SocSciBot3 are analysed by type, to track down the nature of departmental interlinks. The classification scheme used in this project modelled on the target page 'Content' and 'Genre' of Harris et al. (Harries et al., 2004), which is one of the closest previously published studies. The 7 categories are listed below.

- Department homepage. The main page of a department contains the department's name, has its own domain name or directory, and has links to other information about the department.
- Research group homepage. The main page of a research group, centre or laboratory.
- Academic staff member's homepage. The main page of an academic staff member. A homepage of a PhD student is included in this category.
- Resources. A page contains some useful information to the source department, and is not one of the homepages mentioned above. A variety of different contents are included.
- Recreation. A page that only contains recreational information, such as hobbies, sports clubs etc.
- Disappeared. A page that no longer exists when visited. With regard to redirected page, if the page still point to the same position, the page is regarded as still existing, otherwise as disappeared.
- Others. The page does not fit any of the other categories. For example, an undergraduate student's homepage; a page that is password protected, or is not allowed to be viewed for whatever reason.

The number of departmental interlinks is sparse in number compared with university interlinks (Harries et al., 2004), thus it is not plausible to extract as many randomised departmental links from each department (many departments may not link to any other peers at all) as Wilkinson et al. (2003) did for university interlinks. In this project, all target pages are visited to ascertain whether they still exist (many old domain names are included in the project), and to what categories they belong. The approach in Bar-Ilan (2004a) is the most detailed classification scheme of the kind, which classified the source and target page contents together with the link context. However, it is not feasible in this project, as a whole population of target pages will be studied. Nevertheless, if a department is found to attract an exceptionally large number of links from other departments, the source pages are visited to trace the reason. This is a compromise approach with regard to the simplicity of target page identification.

5.5 Computing Different Link Metric Values

This section explains how different values with respect to the four aspects, which are introduced in section 4.2.3, are computed. Each value will be compared along country and disciplinary lines to seek link patterns.

1. General web uses
   i. The median value of web pages (from SocSciBot).
   ii. The median value of external inlinks from the rest of the Web (from AltaVista).
   iii. The median value of external outlinks (from SocSciBot).
   iv. The proportion of national inlinks is calculated as follows:

$$\frac{CI}{TI}$$

- CI: The number of external inlinks from the country domain to a set of departments (from AltaVista).
- TI: The number of external inlinks from the whole Web to the set of departments (from AltaVista).

2. National peer interlinks.
   i. The median value of file ADM inlinks, which are counted through SocSciBot amongst same set of departments, is used to represent the national peer inlinks.
   ii. The interconnection rate for each set of departments is calculated as follows:

$$\frac{P}{TP}$$

- P: The pairs of departments have at least one link from one to another
- TP: Total number of possible pairs among a set of departments

The numerator in fact is the sum of the inlinks received by a set of departments at the department ADM level. The denominator is the total number of possible pairs that may have links from one to another. For example, with regard to the 40 physics departments studied in the UK, the total number of possible pairs may have links from one to another is $40 \times (40\text{-}1)=1560$

3. International peer interlinks.

The number of links between two sets of departments in a same discipline from two different countries is collected through AltaVista. The Link Propensity and adapted mean international peer inlinks for each set of departments are then calculated, to illustrate how the departments attract international peer inlinks.

i. Link propensities. The link propensity between two countries' same type of departments can be calculated as follows:

$$\frac{DL}{PN}$$

- DL: The links from source departments in one country to the same type of target departments in another country.
- PN: The product of number of staff members (or number of web pages) from source and target

The product of staff members:

the number of source staff members × the number of target staff members

The product of number of web pages:

the number of source web pages × the number of target web pages

ii. Adapted mean international peer inlinks.

Each country has a different number of departments in each discipline. The three types of Australian departments all have the smallest numbers amongst the three countries. When counting links from the UK and Canada to Australia, the number of source departments (the departments in the UK plus those in Canada) is larger than when counting links to the UK or Canada. In order to remove this size effect, the number of inlinks received by each set of departments is divided by the source departments in the two other countries. The number of international peer inlinks for each set of departments is thus adapted to be the number of inlinks expected to be received from one source department. The mean value for each set of departments is then calculated by dividing the number of the adapted international peer inlinks by the number of departments from each set. The two steps can be combined as one.

The adapted mean international peer (AIP) inlinks is thus calculated as follows:

$$AIP = PI/PD$$

- PI: The inlinks received from peers in two other countries.
- PD: Product of number of departments from source and target
  The product of number of departments:
  the number of source departments × the number of target departments

4. Interactions with different top level domains.

i. The proportion of outlinks to a top level domain from a set of departments will be calculated as follows:

$$\frac{EO}{TO}$$

- EO: Number of external outlinks made to a top level domain from a set of departments (from SocSciBot).
- TO: Total number of external outlinks made by a set of departments (from ScoSciBot).

ii. The proportion of external inlinks from a top level domain will be calculated as follows:

$$\frac{EI}{TI}$$

- EI: Number of external inlinks received from a top level domain (from AltaVista)
- TI: Number of external inlinks from the whole Web (from AltaVista)

5.6 Summary

In order to reach the two main objectives of this study, the correlation tests, target page classification and computing of different values with regard to four aspects, need to be carried out. First of all, relevant data needs to be collected. The departmental link data is collected both through SocSciBot and AltaVista. SocSciBot is used when the link data is within a department or a set of departments' web sites. AltaVista is chosen when the link data is from large web areas, such as the whole Web or different top level domains. Some non-web data, such as, departmental research measures and number of academic staff members are also collected in order to carry out the experiment. A semi-automatic technique is introduced to count citations for Canadian and UK departments from the ISI's Web of Science.

For each of the three countries, universities' link structures are crawled first through SocSciBot. Then a department's link structure is extracted from its parent university's link structure. Existing universities' link structure can be used without repeat crawling on the Web. The departmental link structure includes more link pages than crawling each department's web site separately. A set of software is used to count relevant numbers for this study. They are:

1. Number of web pages each department publishes
2. Number of inlinks and outlinks with four different ADMs from national peers
3. Number of external outlinks made from each department
4. Number of outlinks to a top level domain from a set of departments

AltaVista collects:

1. Number of external inlinks from the whole Web
2. Number of external inlinks from different top level domains
3. Number of web pages for each department

In order to count citations from the ISI's Web of Science, this section explains how to:

1. Collect relevant ISI records for each set of departments
2. Save the ISI records in a plain text file
3. Count citations through matching departments' postcodes
4. Validate the citation counts

This section designs a set of correlation with regard to SocSciBot and AltaVista link data. It also explains how to carry out the target page type classification scheme in detail. Finally, the section explains how each value with regard to the four aspects is calculated in detail.

6. Validation Tests

Both the correlations between link counts and research measures, and the results of the target page type classification are reported in this chapter.

6 .1 Results for the Correlation Tests

Tables 6.1 to 6.60 display the correlation coefficient values between link and research measures for the physics, chemistry and biology departments in Australia, Canada and the UK. The '(staff)' indicates that the denominators for the 'WIFs' or WUFs' are the numbers of staff members, otherwise the denominators are the numbers of web pages, directories, domains or departments where appropriate. In order to save space, only four diagrams are illustrated to show the relationship between link and research measures for each set of departments (figures 6.1 to 6.36). The four types of figures are plotted between inlinks at the file level counted by the SocSciBot3 and research productivities; between the external inlinks counted by AltaVista and research productivities; between file WIFs (staff) and research averages; between external WIFs (staff) and research averages. Section 6.1.4 reports the correlation coefficient values between citation counts and research measures for Canadian and UK departments, as shown in tables 6.61 to 6.64.

6.1.1 Correlation Coefficient Values for Australian Departments

6.1.1.1 Correlation Coefficient Values for the Physics Departments

Tables 6.1 and 6.2 display the Pearson and Spearman correlations between the number of links to or from the physics departments with different ADMs and number of citations; and between WIFs (staff) or WUFs (staff) and citations per staff member or CPP if the denominators of the WIFs or WUFs are the numbers of web pages, directories, domains or departments. Tables 6.3 and 6.4 show the Pearson and Spearman correlations between the number of inlinks from different domains and citations; and between WIFs (staff) and citations per staff member or CPP if the denominators of the WIFs are the numbers of web pages. The correlation coefficient values between the number of links and citations, between the WIFs (staff) or WUFs (staff) and citations per staff member are all significant at the 1% level, see tables 6.1 to 6.4. However, hardly any significant correlation coefficient values are found between the WIFs or WUFs (with the numbers of web pages, directories, domains or departments as denominators) and CPP.

Strong linear trends are shown in figures 6.1 to 6.4. This is the reflection of the large correlation coefficient values for the Australian physics departments. The reason may be that the links made to or from the set of departments are larger (see tables 6.69, 7.2, 7.3 and 7.6) and more even compared with other sets of departments. In table 7.5, the Australian physics departments interconnect much better than the chemistry and biology departments.

Table 6.1 Pearson correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=22)

Link measures	File	Directory	Domain	Department
Inlinks	0.847**	0.881**	0.711**	0.575**
Outlinks	0.841**	0.860**	0.856**	0.786**
WIFs (staff)	0.839**	0.843**	0.639**	0.571**
WUFs (staff)	0.787**	0.781**	0.675**	0.701**
WIFs	0.011	-0.079	0.217	0.457
WUFs	-0.041	0.094	0.426	0.548*

Table 6.2 Spearman correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=22)

Link measures	File	Directory	Domain	Department
Inlinks	0.890**	0.901**	0.884**	0.836**
Outlinks	0.818**	0.821**	0.829**	0.849**
WIFs (staff)	0.842**	0.860**	0.768**	0.650**
WUFs (staff)	0.780**	0.783**	0.767**	0.794**
WIFs	0.083	0.087	0.378	0.508*
WUFs	0.261	0.182	0.540*	0.588*

Table 6.3 Pearson correlations between the link (from AltaVista) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=22 for staff members as denominators, n=24 for the rest)

Source Domain	inlinks and citations	WIFs (staff) and citations per staff member	WIFs and CPP
External	0.900**	0.927**	-0.064
com	0.798**	0.854**	0.010
edu	0.790**	0.720**	-0.069
au	0.850**	0.903**	-0.104
org	0.700**	0.854**	0.033
net	0.639**	0.659**	0.083
com.au	0.972**	0.909**	-0.068
edu.au	0.756**	0.838**	-0.104

Table 6.4 Spearman correlations between the link (from AltaVista) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=22 for staff member as denominators, n=24 for the rest)

Source Domain	inlinks and citations	WIFs (staff) and citations per staff member	WIFs and CPP
External	0.887**	0.867**	-0.076
com	0.793**	0.735**	0.070
edu	0.818**	0.830**	-0.121
au	0.902**	0.889**	-0.172
org	0.724**	0.747**	-0.111
net	0.717**	0.739**	0.292
com.au	0.790**	0.842**	-0.132
edu.au	0.878**	0.877**	-0.247

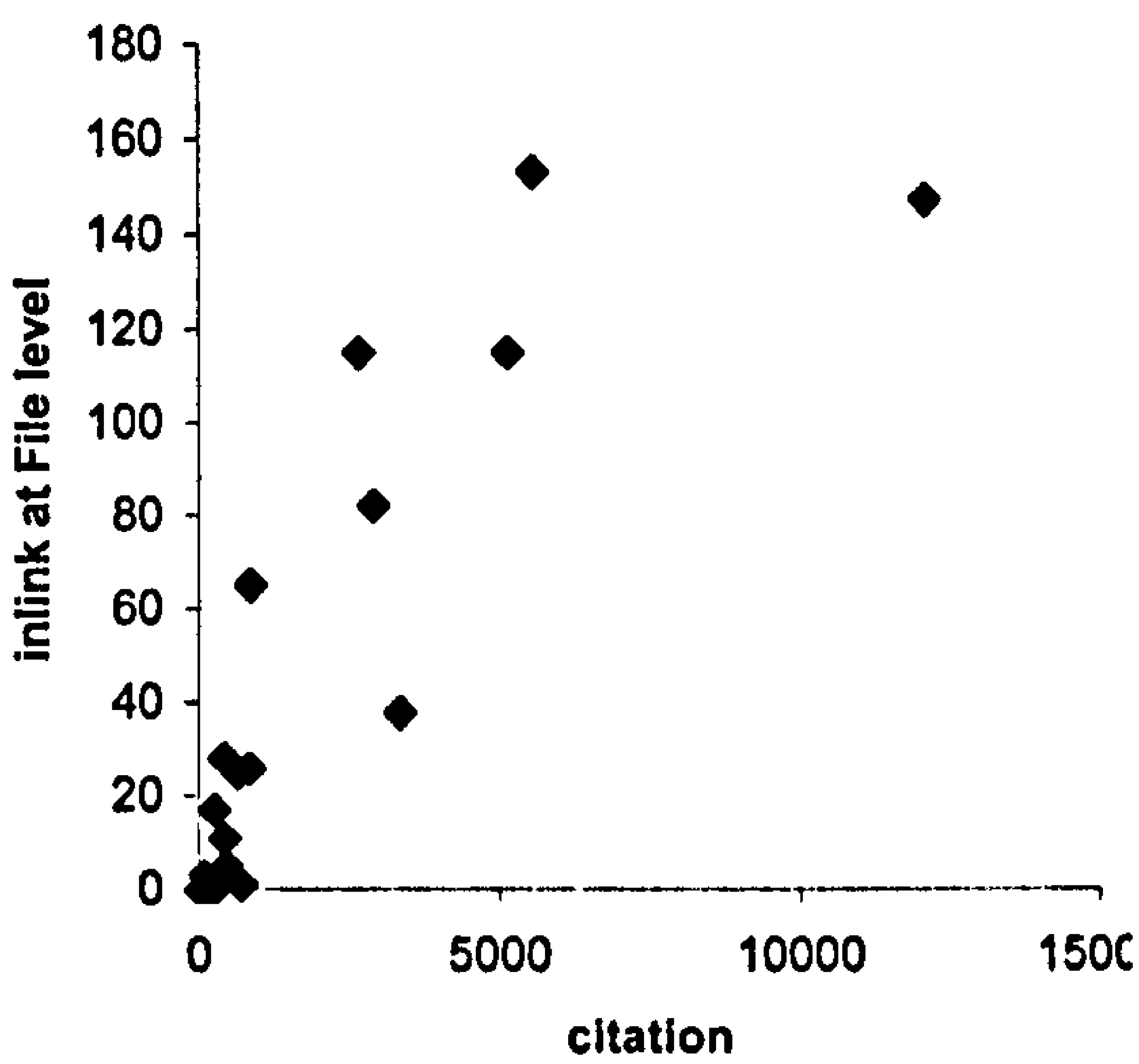


Figure 6.1 File inlinks against citations for Australian physics departments

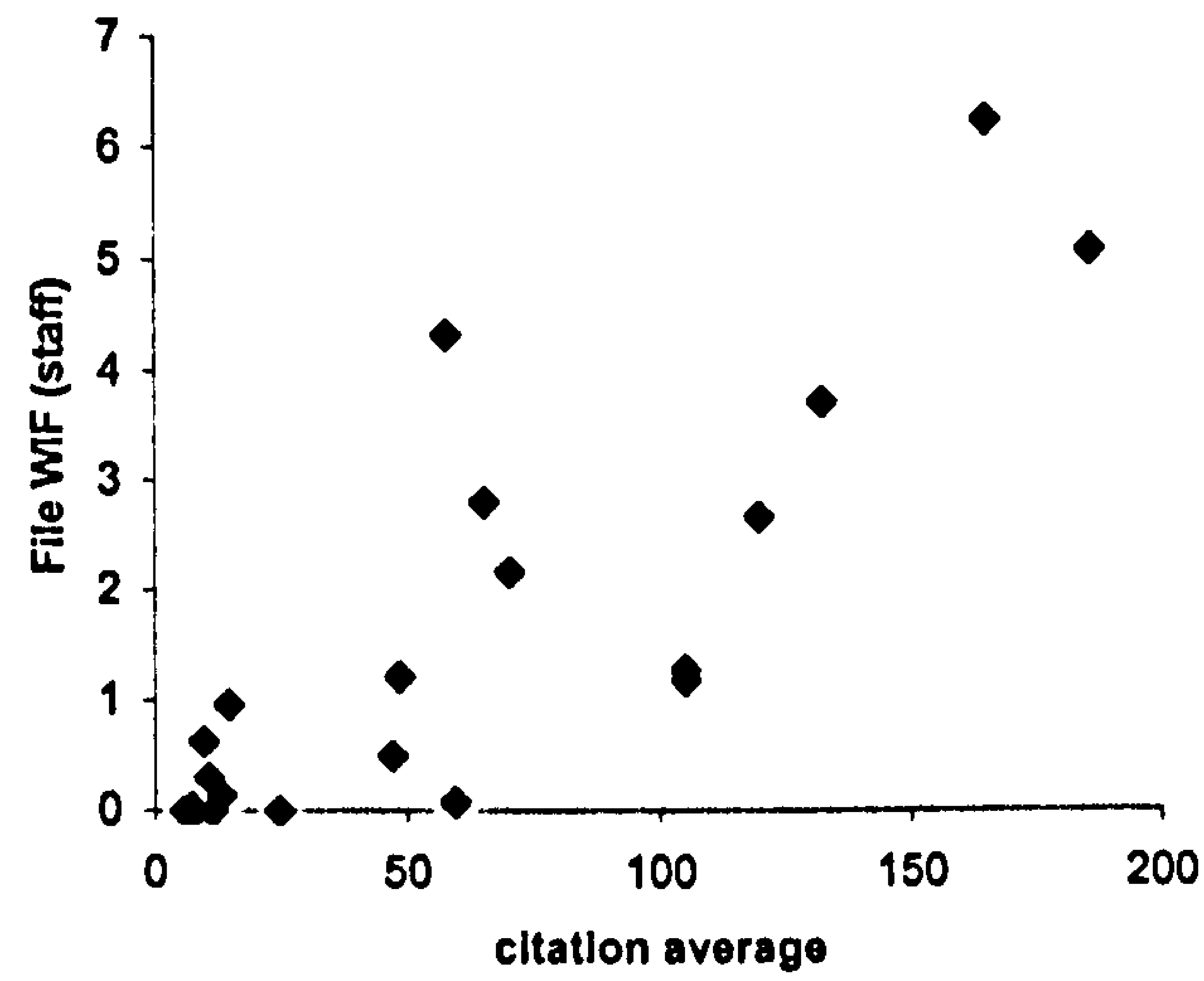


Figure 6.3 File WIFs (staff) against average citations for Australian physics departments

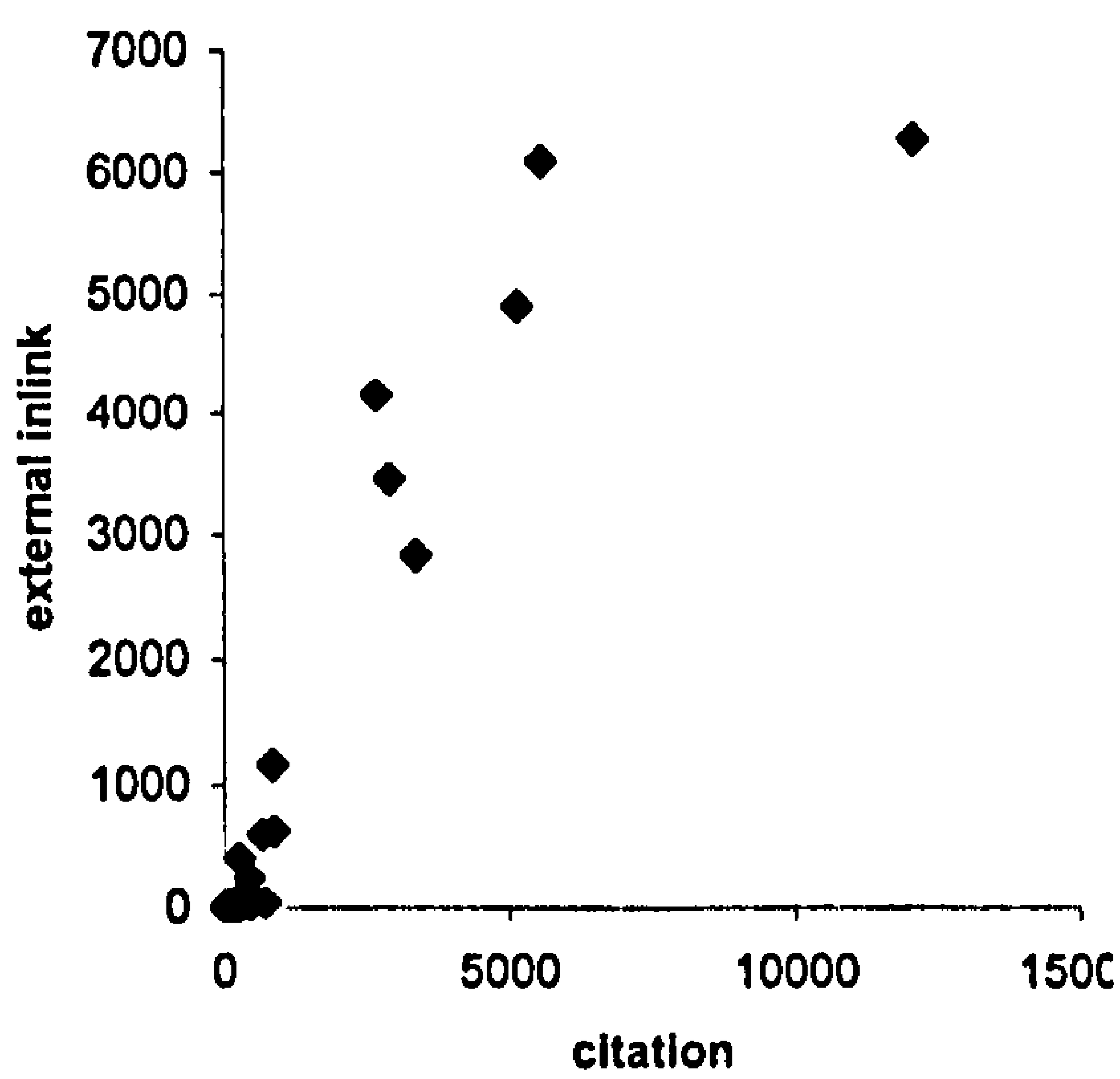


Figure 6.2 External inlinks against citations for Australian physics departments

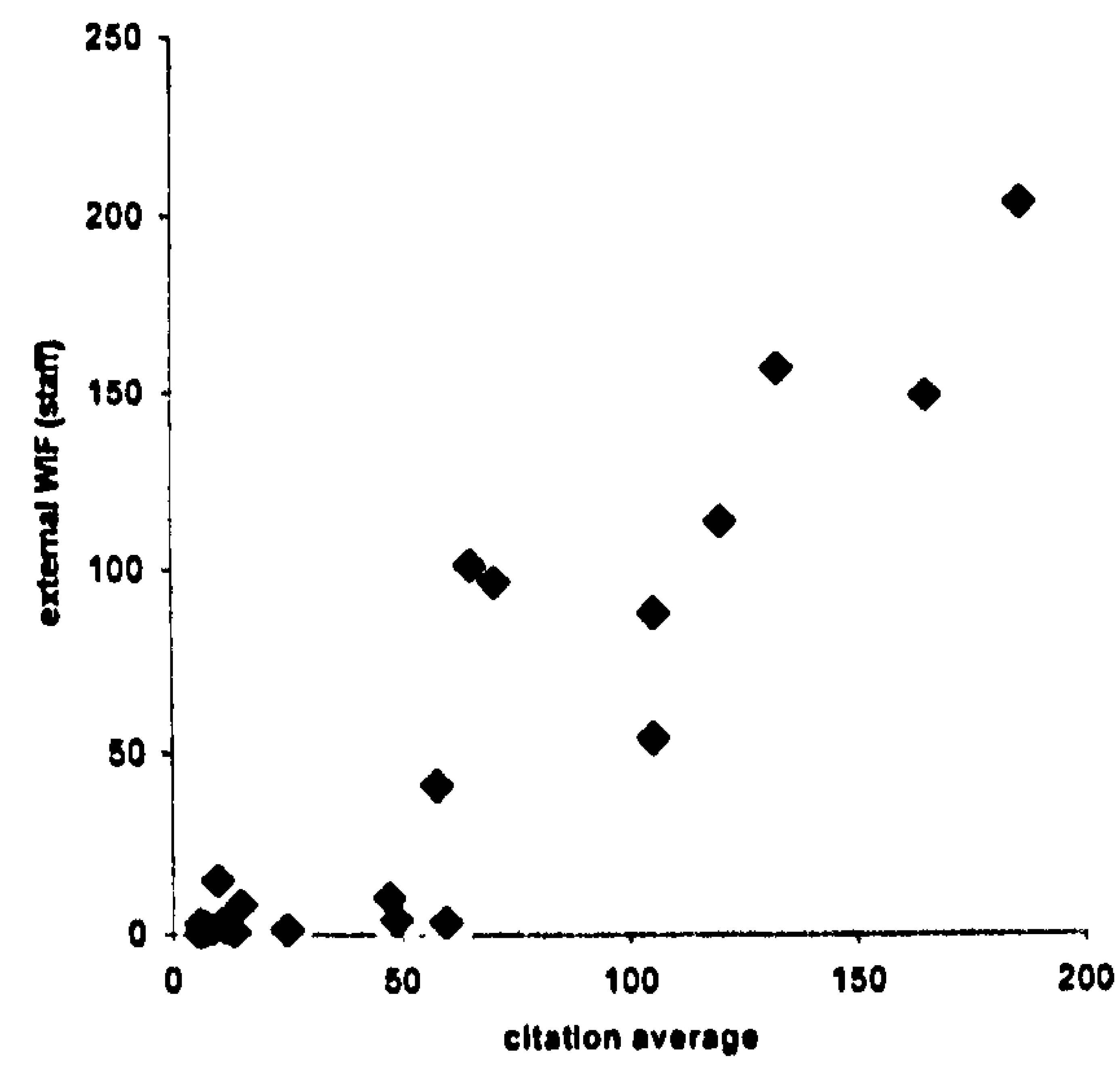


Figure 6.4 External WIFs (staff) against average citations for Australian physics departments

6.1.1.2 Correlation Coefficient Values for the Chemistry Departments

Tables 6.5 and 6.6 display the Pearson and Spearman correlations between the number of links to or from the chemistry departments (with different ADMs) and citations; and between WIFs (staff) or WUFs (staff) and the number of citations per staff member or CPP when the denominators for the WIFs or WUFs are the number of web pages, directories, domains and departments. Tables 6.7 and 6.8 show the Pearson and Spearman correlations between the number of inlinks from different domains and citations; and between WIFs (staff) and citations per staff member or CPP if the denominators of the WIFs are the numbers of web pages.

The correlation coefficient values between the number of links with different ADMs and citations are all significant at the 1% level, see tables 6.5 and 6.6. The correlation coefficient values between the WIFs (staff) or WUFs (staff) and citations per staff are found more significant than those between the WIFs or WUFs (with the numbers of web pages, directories, domains and departments as denominators) and CPP. The Spearman values are relatively larger than those Pearson ones.

Figures 6.5 to 6.8 do not show strong linear trends as those in figures 6.1 to 6.4 of the physics departments. The set of departments does not interlink as well as the physics departments, as shown in table 7.5. The correlation coefficient values for the chemistry departments are smaller than those of the physics departments.

Table 6.5 Pearson correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=25)

Link measures	File	Directory	Domain	Department
Inlinks	0.558**	0.605**	0.649**	0.699**
Outlinks	0.697**	0.717**	0.739**	0.754**
WIFs (staff)	0.294	0.369	0.335	0.498*
WUFs (staff)	0.100	0.333	0.291	0.277
WIFs	-0.071	0.509*	0.396	0.545*
WUFs	-0.015	0.242	0.300	0.407

Table 6.6 Spearman correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=25)

Link measures	File	Directory	Domain	Department
Inlinks	0.588**	0.622**	0.620**	0.667**
Outlinks	0.564**	0.565**	0.554**	0.554**
WIFs (staff)	0.455*	0.529**	0.550**	0.565**
WUFs (staff)	0.382	0.421*	0.398*	0.398*
WIFs	0.121	0.473	0.515*	0.578*
WUFs	0.195	0.335	0.387	0.458

Table 6.7 Pearson correlations between the link (from AltaVista) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=25)

Source Domain	inlinks and citations	WIFs (staff) and citations per staff member	WIFs and CPP
External	0.410*	0.322	0.252
com	0.743**	0.625**	0.215
edu	0.751**	0.521**	0.239
au	0.247	0.265	0.254
org	0.710**	0.358	0.200
net	0.530**	0.408*	0.253
com.au	0.781**	0.519**	0.031
edu.au	0.221	0.252	0.255

Table 6.8 Spearman correlations between the link (from AltaVista) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=25)

Source Domain	inlinks and citations	WIFs (staff) and citations per staff member	WIFs and CPP
External	0.670**	0.597**	-0.038
com	0.563**	0.520**	0.021
edu	0.747**	0.540**	-0.121
au	0.658**	0.632**	0.138
org	0.464*	0.354	0.099
net	0.671**	0.478*	0.202
com.au	0.705**	0.499*	0.049
edu.au	0.615**	0.601**	0.254

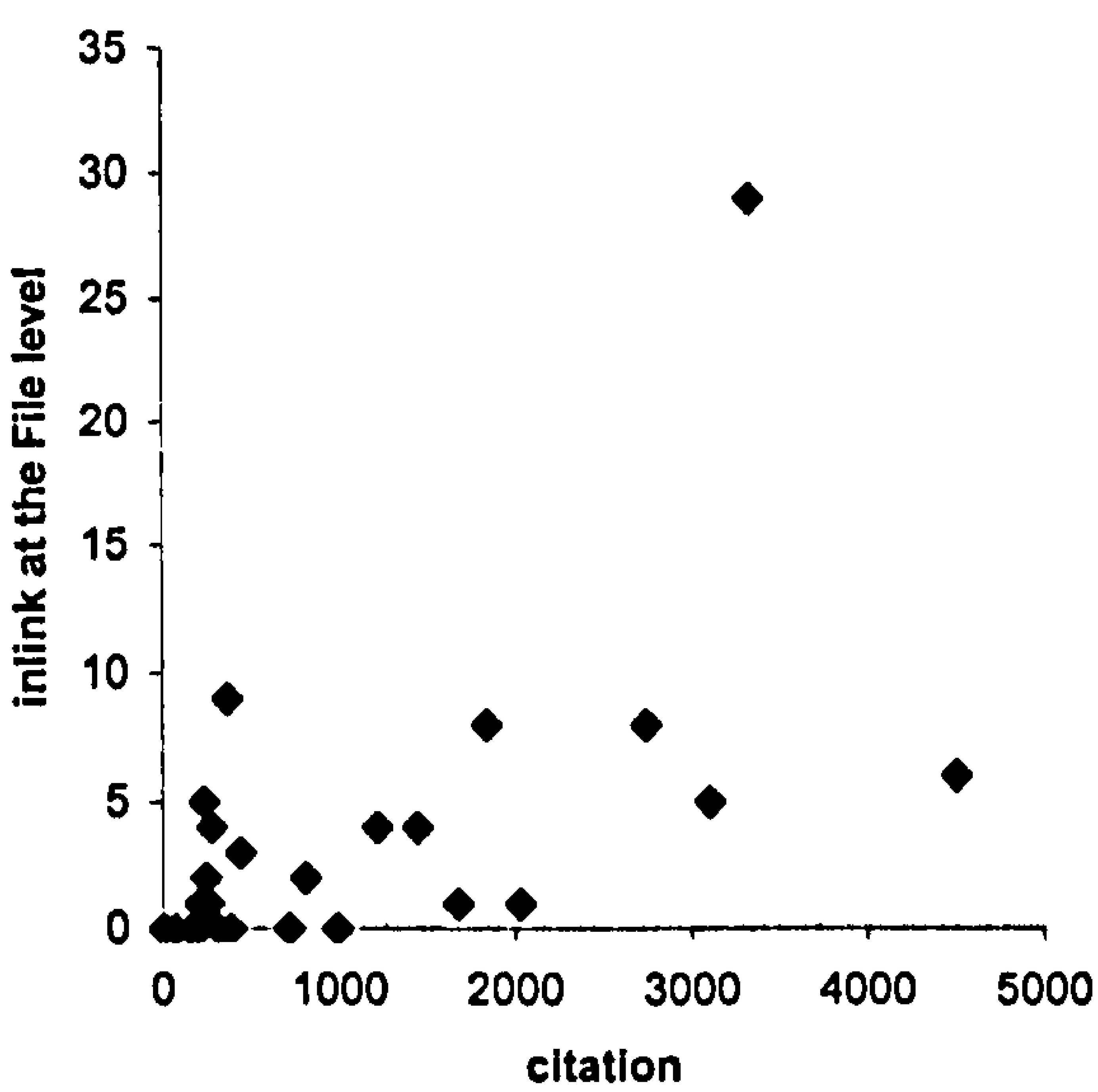


Figure 6.5 File inlinks against citations for Australian chemistry departments

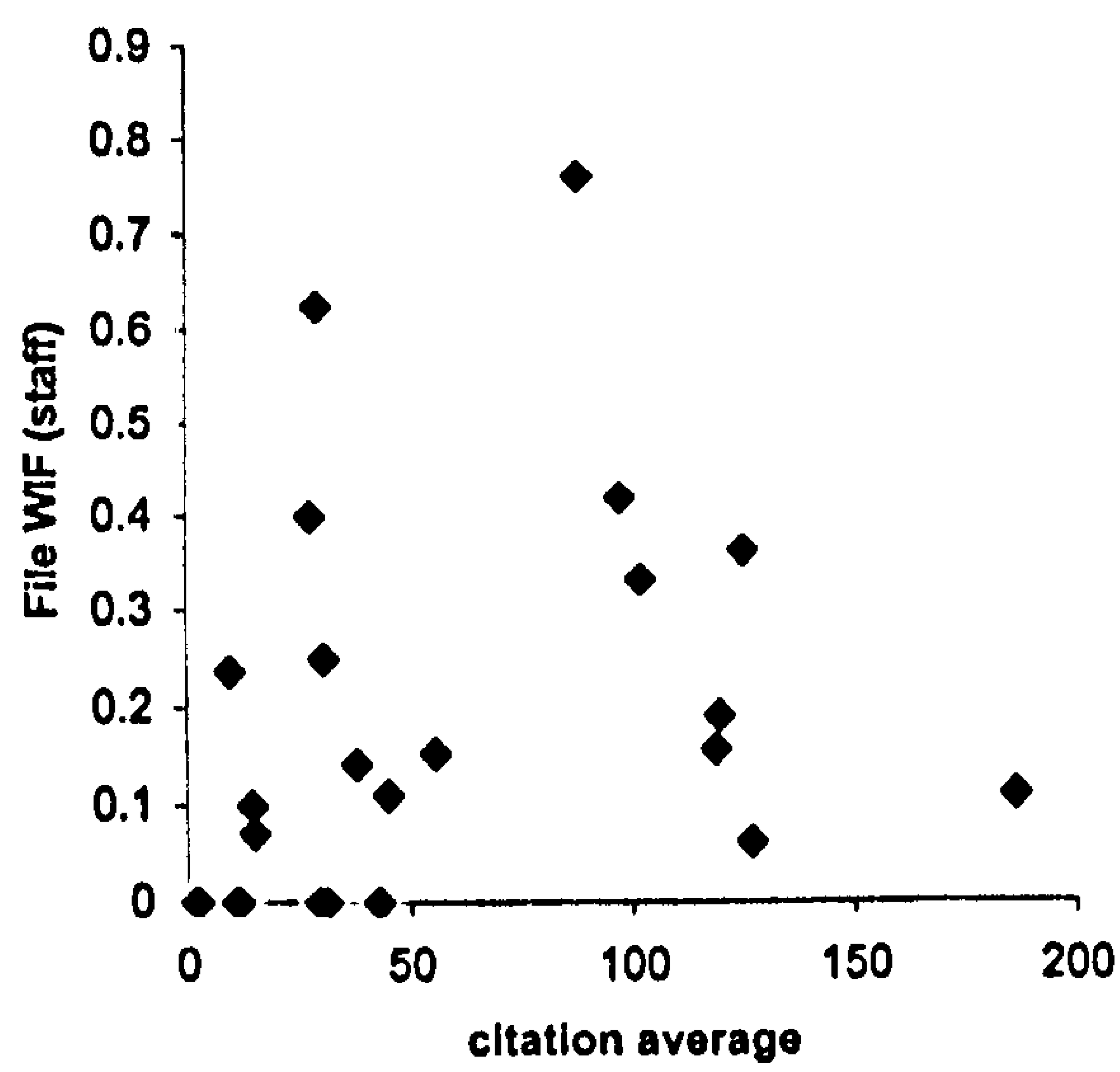


Figure 6.7 File WIFs (staff) against average citations for Australian chemistry departments

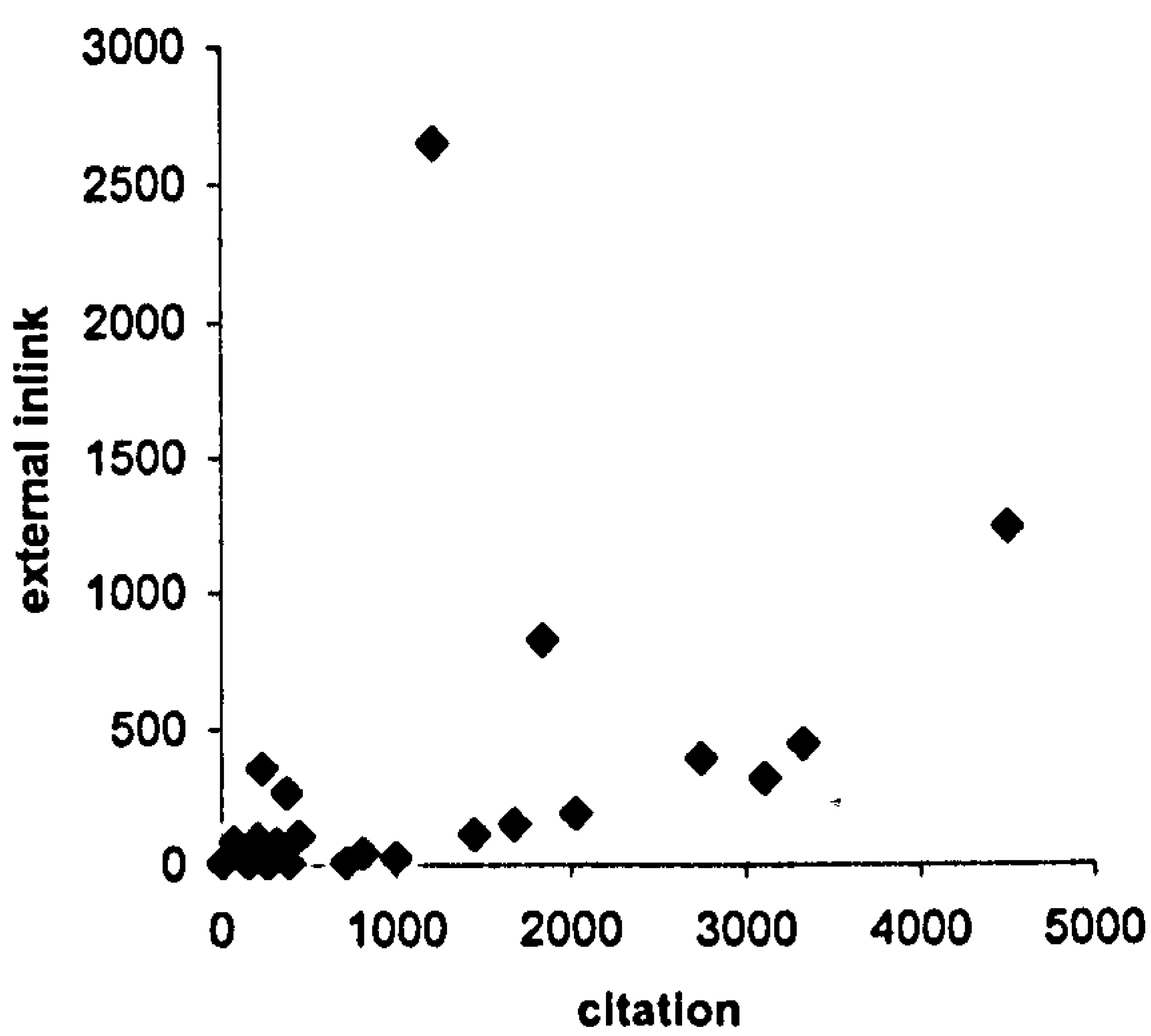


Figure 6.6 External inlinks against citations for Australian chemistry departments

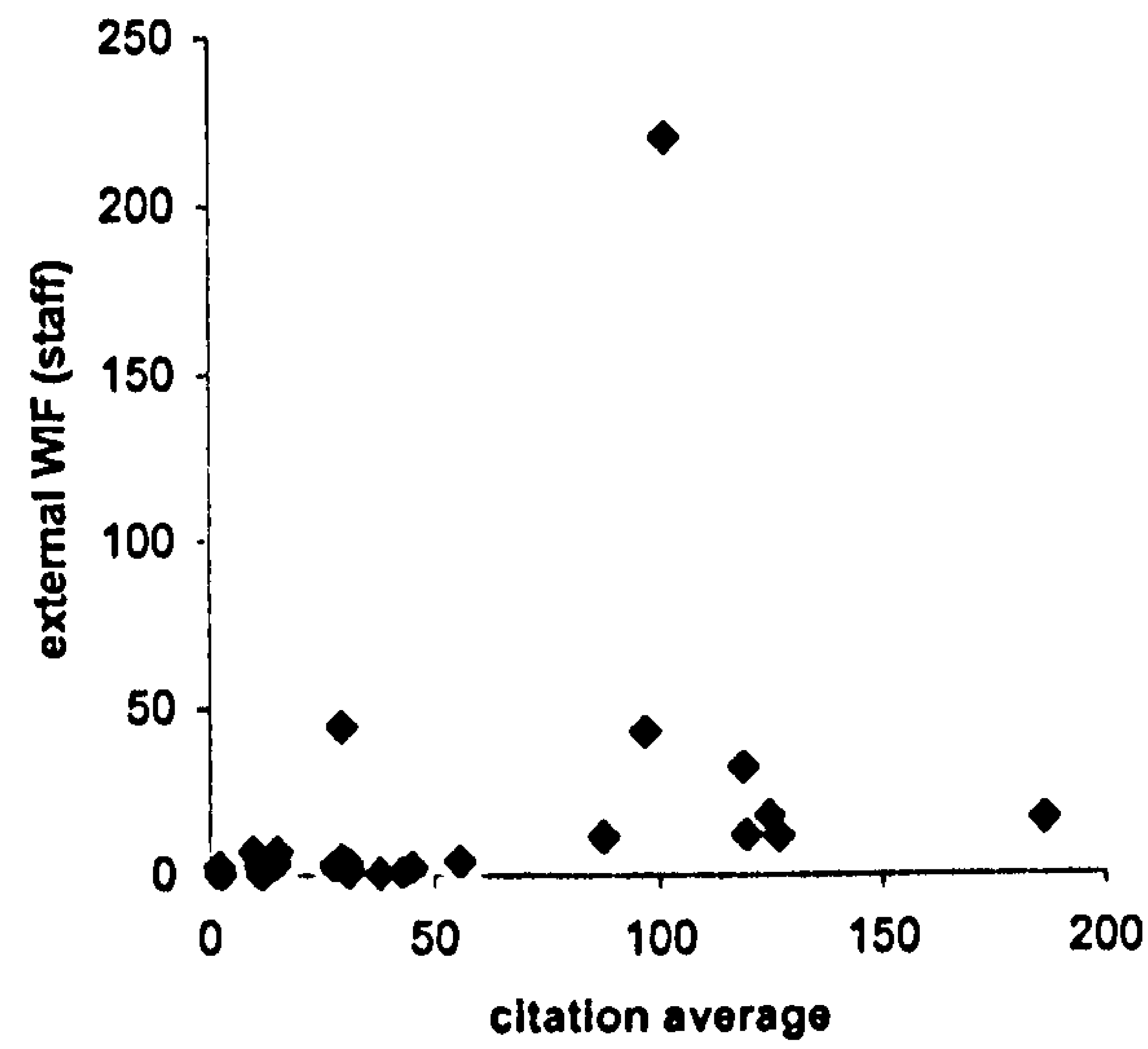


Figure 6.8 External WIFs (staff) against average citations for Australian chemistry departments

6.1.1.3 Correlation Coefficient Values for the Biology Departments

Tables 6.9 and 6.10 display the Pearson and Spearman correlations between the number of links to or from the biology departments with different ADMs and citations; and between WIFs (staff) or WUFs (staff) and citations per staff member or CPP if the denominators for the WIFs or WUFs are the numbers of web pages, directories, domains and departments. Tables 6.11 and 6.12 list the Pearson and Spearman correlation coefficient values between the number of inlinks from different domains and citations; and between WIFs (staff) and citations per staff member or CPP if the denominators of the WIFs are the numbers of web pages.

The correlation coefficient values between the number of links and citations are all significant at the 1% level, as shown in tables 6.9 to 6.12. Pearson correlation coefficient values between the WUFs (staff) with different ADMs and citations per staff member are significant at the 1% level in table 6.9, while Spearman correlation coefficient values are significant at the 5% level in table 6.10. It is a surprise that no significant values are found between the WIFs (staff) and citations per staff member both in tables 6.9 and 6.10. The small amount of links between the Australian biology departments may be the reason. As shown in table 6.69, there are altogether 90 links among the Australian biology departments, while the median number of inlinks from the peer departments in the same set is only 1, as show in table 7.6. There are some significant values between the WIFs (staff) and citations per staff member in table 6.12. However, no significant values are found between the WIFs or WUFs (with numbers of web pages, directories, domains and departments as denominators) and CPP in tables 6.9 to 6.12.

The linear trends in figures 6.9 to 6.12 again are not as apparent as those for the physics departments in figures 6.1 to 6.4. However, figures 6.9 and 6.10 illustrate better linear trends than those in figures 6.11 and 6.12.

Table 6.9 Pearson correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=26)

Link measures	File	Directory	Domain	Department
Inlinks	0.707**	0.666**	0.746**	0.684**
Outlinks	0.581**	0.597**	0.649**	0.641**
WIFs (staff)	0.247	0.206	0.160	0.165
WUFs (staff)	0.708**	0.740**	0.742**	0.726**
WIFs	-0.209	-0.055	0.108	0.294
WUFs	0.139	-0.035	0.070	0.247

Table 6.10 Spearman correlations between the link (from SocSciBot) and research measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=26)

Link measures	File	Directory	Domain	Department
Inlinks	0.586**	0.585**	0.590**	0.589**
Outlinks	0.660**	0.670**	0.704**	0.696**
WIFs (staff)	0.270	0.274	0.237	0.239
WUFs (staff)	0.429*	0.458*	0.445*	0.442*
WIFs	0.327	0.088	0.244	0.454
WUFs	0.365	0.052	0.216	0.430

Table 6.11 Pearson correlations between the link (from AltaVista) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=26, n=25 for WIFs with numbers of web pages as denominators)

Source Domain	inlinks and citations	WIFs (staff) and citations per staff member	WIFs and CPP
External	0.573**	0.227	-0.099
com	0.540**	0.096	-0.144
edu	0.421*	0.126	0.217
au	0.814**	0.258	-0.109
org	0.671**	0.208	0.399*
net	0.511**	0.040	0.085
com.au	0.592**	0.031	-0.326
edu.au	0.781**	0.264	-0.103

Table 6.12 Spearman correlations between the link (from AltaVista) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=26, n=25 for WIFs with numbers of web pages as denominators)

Source Domain	inlinks and citations	WIFs (staff) and citations per staff member	WIFs and CPP
External	0.678**	0.480*	0.107
com	0.638**	0.521**	0.117
edu	0.562**	0.325	0.332
au	0.678**	0.482*	0.152
org	0.590**	0.354	0.654**
net	0.606**	0.198	0.344
com.au	0.563**	0.183	0.127
edu.au	0.608**	0.533**	0.149

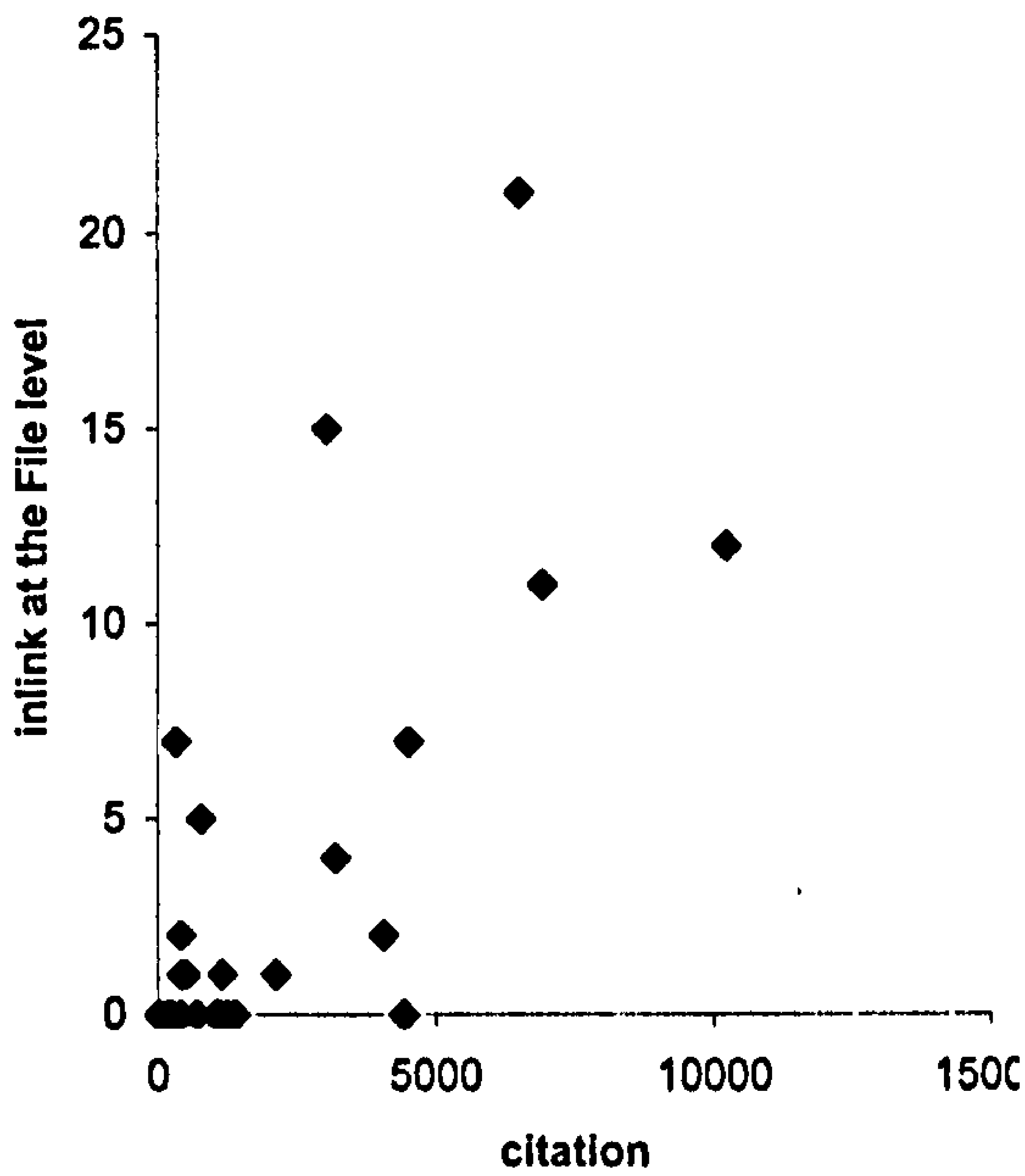


Figure 6.9 File inlinks against citations for Australian biology departments

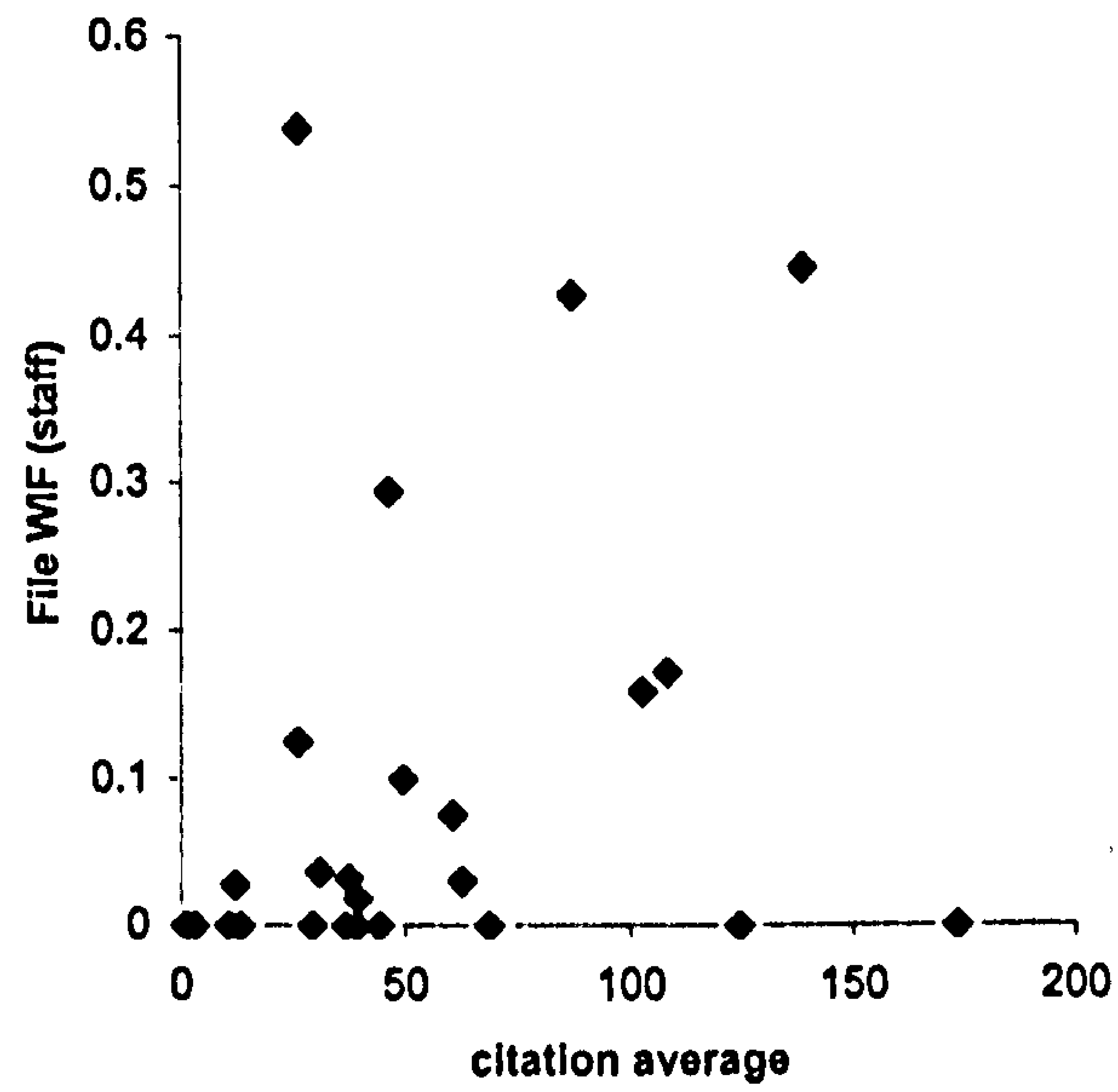


Figure 6.11 File WIFs (staff) against average citations for Australian biology departments

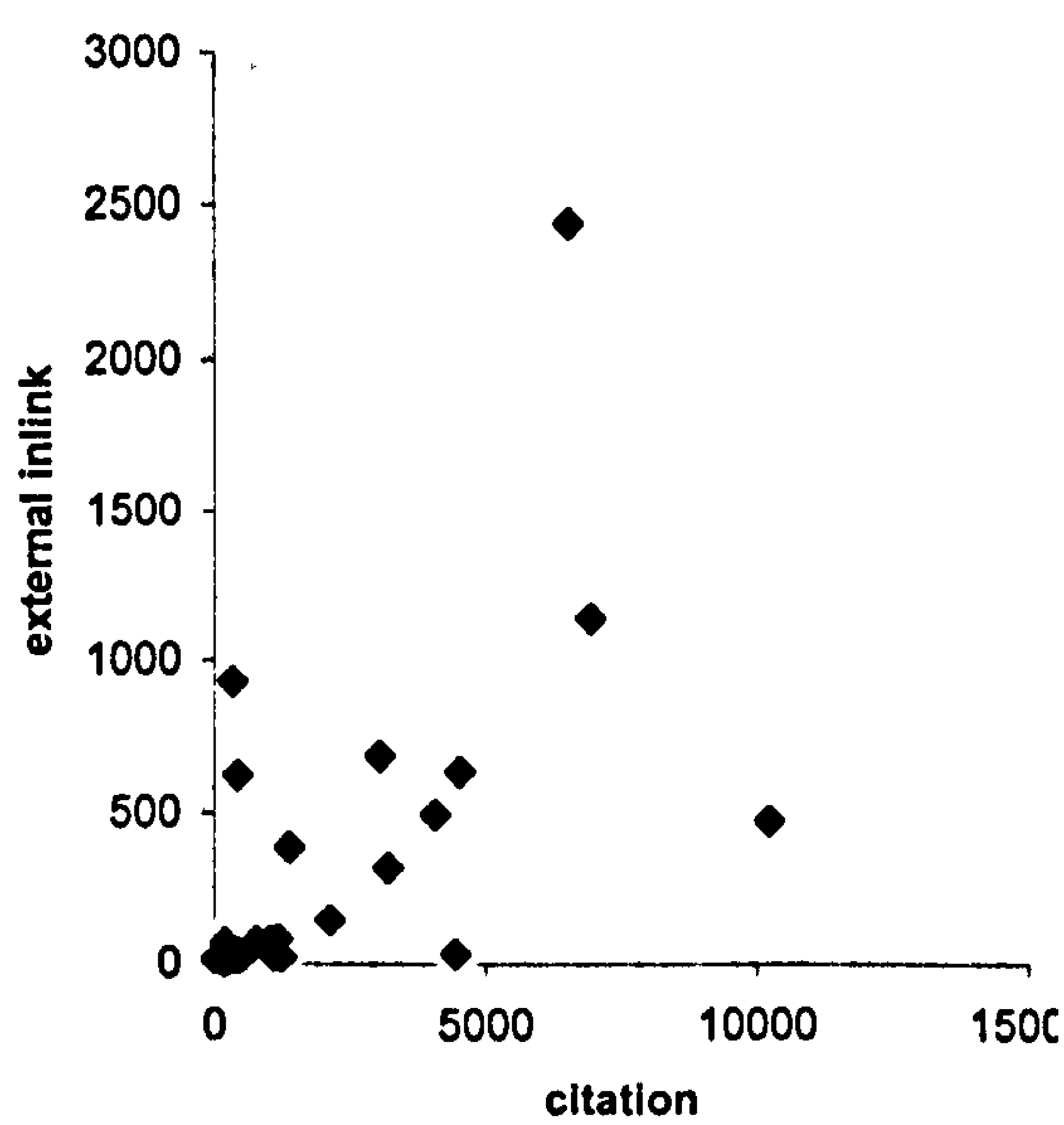


Figure 6.10 External inlinks againt citations for Australian biology departments

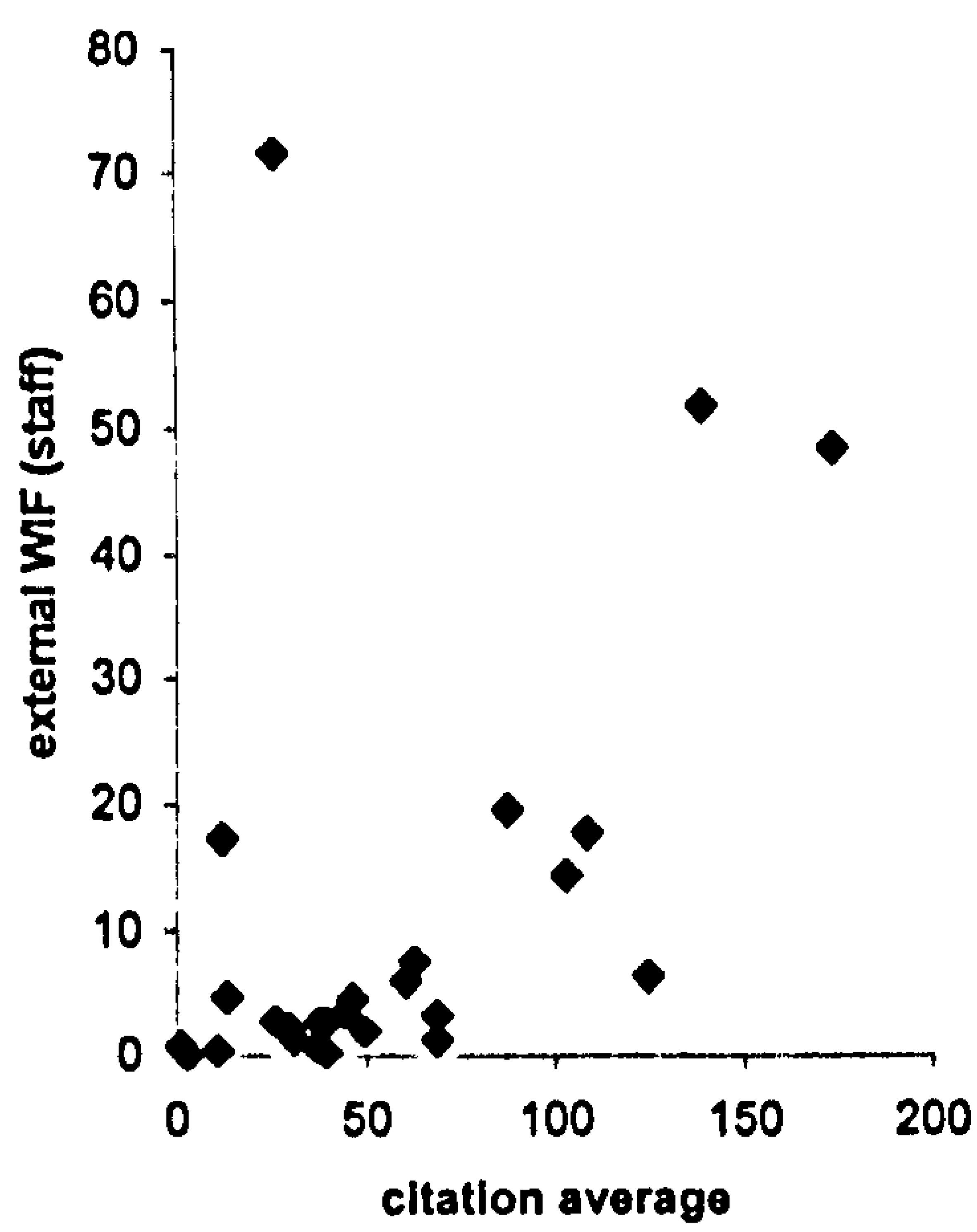


Figure 6.12 External WIFs (staff) against average citations for Australian biology departments

6.1.2 Correlation Coefficient Values for Canadian Departments

6.1.2.1 Correlation Coefficient Values for the Physics Departments

Tables 6.13, 6.14, 6.17 and 6.18 list the Pearson and Spearman correlations between the number of links to or from the physics departments with different ADMs and citations or research grants; and between WIFs or WUFs and citation or research grants averages. Tables 6.15, 6.16, 6.19 and 6.20 illustrate the Pearson and Spearman correlations between the inlinks from different domains and citations or research grants; and between WIFs and citation or research grants averages.

The correlation coefficient values between the links and citations or research grants are nearly all significant at the 1% level in tables 6.13 to 6.20. There are several significant correlation coefficient values between the WIFs and citation or research grants averages in tables 6.13 to 6.20. The value of the department WIFs is in fact the number of inlinks at the department level, because the denominator is 1 for each department. The significant values can also be regarded as the department inlinks correlating significantly with citation or research grants averages, although the values are smaller than with the citation counts or research grants. More significant values are found between different domain WIFs (staff) and research grants per staff member than those between the WIFs (with number of web pages as denominators) and research grants per staff member, as illustrated in tables 6.15, 6.16, 6.19 and 6.20.

The linear trend between the inlinks and research grants, as illustrated in figures 6.13 and 6.15, are more apparent than those between the WIFs and research averages, as illustrated in figures 6.14 and 6.16.

Table 6.13 Pearson correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=37, n=33 for WIFs and WUFs)

Link measures	File	Directory	Domain	Department
Inlinks	0.600**	0.650**	0.695**	0.715**
Outlinks	0.630**	0.581**	0.551**	0.356*
WIFs (staff)	0.144	0.104	0.230	0.001
WUFs (staff)	0.341	0.340	0.281	0.142
WIFs	-0.166	-0.155	0.256	0.650**
WUFs	0.127	0.142	0.125	0.435*

Table 6.14 Spearman correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=37, n=33 for WIFs and WUFs)

Link measures	File	Directory	Domain	Department
Inlinks	0.670**	0.662**	0.769**	0.733**
Outlinks	0.520**	0.515**	0.444**	0.419**
WIFs (staff)	0.171	0.146	0.208	0.078
WUFs (staff)	0.336	0.301	0.242	0.055
WIFs	0.014	0.191	0.234	0.523**
WUFs	0.092	0.254	0.247	0.504**

Table 6.15 Pearson correlations between the link (from AltaVista) and research (citations) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=37)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.758**	0.398*	0.301
com	0.713**	0.215	0.280
edu	0.690**	0.394*	0.304
ca	0.702**	0.307	0.091
org	0.677**	0.365*	0.420**
net	0.733**	0.242	0.343*

Table 6.16 Spearman correlations between the link (from AltaVista) and research (citations) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=37)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.688**	0.428**	0.468**
com	0.623**	0.286	0.404*
edu	0.670**	0.381*	0.360*
ca	0.714**	0.350*	0.203
org	0.680**	0.411*	0.695**
net	0.663**	0.413*	0.509**

Table 6.17 Pearson correlations between the link (from SocSciBot) and research (research grants) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=37)

Link measures	File	Directory	Domain	Department
Inlinks	0.695**	0.731**	0.773**	0.744**
Outlinks	0.624**	0.690**	0.659**	0.583**
WIFs (staff)	0.316	0.335	0.411*	0.158
WUFs (staff)	0.243	0.268	0.162	0.040
WIFs	-0.123	0.012	0.221	0.384*
WUFs	0.075	0.133	0.097	0.376*

Table 6.18 Spearman correlations between the link (from SocSciBot) and research (research grants) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=37)

Link measures	File	Directory	Domain	Department
Inlinks	0.669**	0.675**	0.738**	0.726**
Outlinks	0.594**	0.594**	0.532**	0.516**
WIFs (staff)	0.217	0.234	0.236	0.157
WUFs (staff)	0.354	0.348	0.275	0.189
WIFs	0.208	0.214	0.176	0.408*
WUFs	0.242	0.269	0.238	0.438*

Table 6.19 Pearson correlations between the link (from AltaVista) and research (research grants) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=37)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.766**	0.502**	0.327*
com	0.702**	0.320	0.300
edu	0.778**	0.528**	0.279
ca	0.701**	0.319	0.006
org	0.732**	0.583**	0.173
net	0.585**	0.305	0.273

Table 6.20 Spearman correlations between the link (from AltaVista) and research (research grants) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=37)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.706**	0.431**	0.424**
com	0.668**	0.344*	0.467**
edu	0.719**	0.475**	0.112
ca	0.702**	0.314	0.061
org	0.725**	0.489**	-0.012
net	0.664**	0.384*	0.060

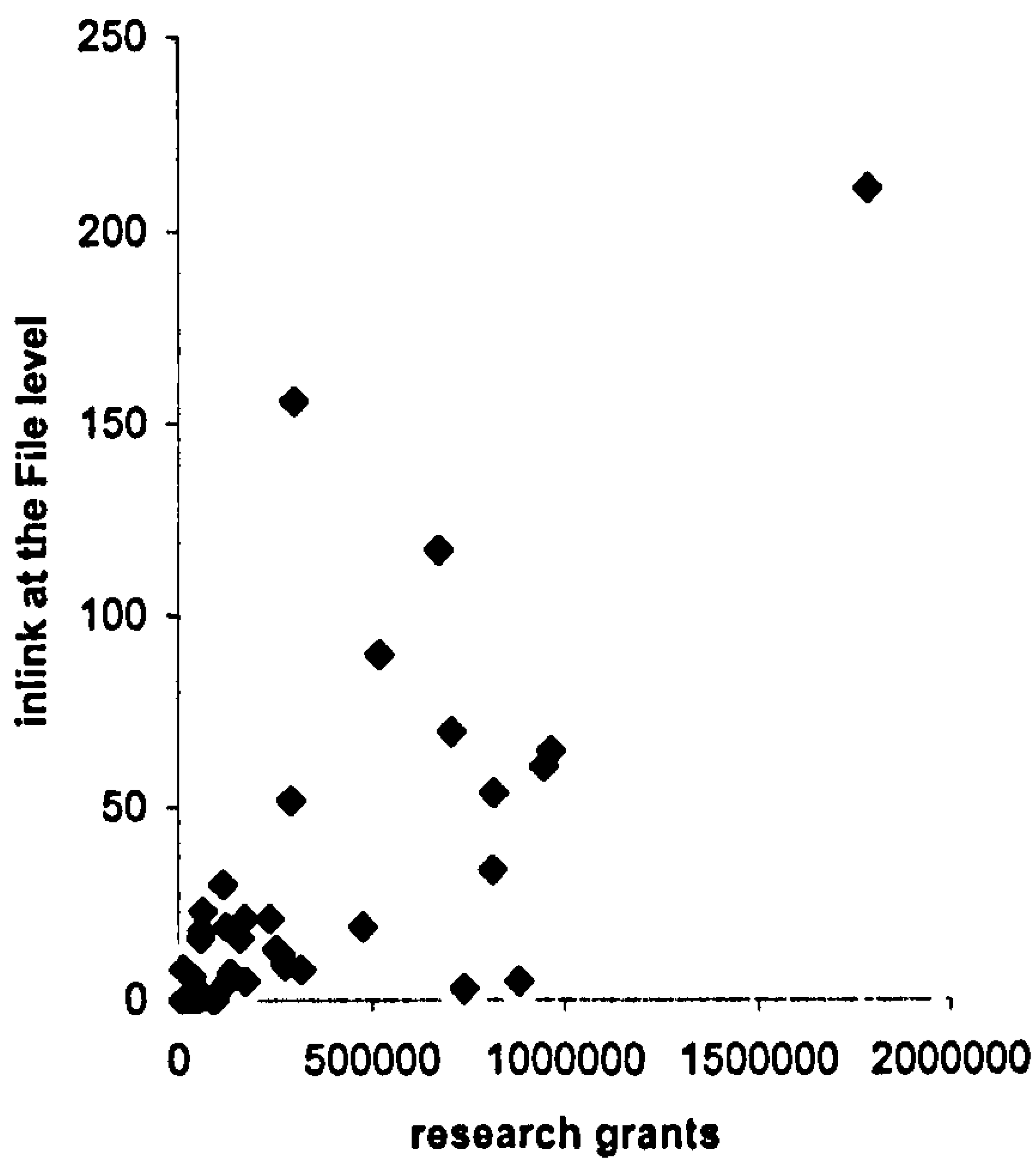


Figure 6.13 File inlinks against research grants for Canadian physics departments

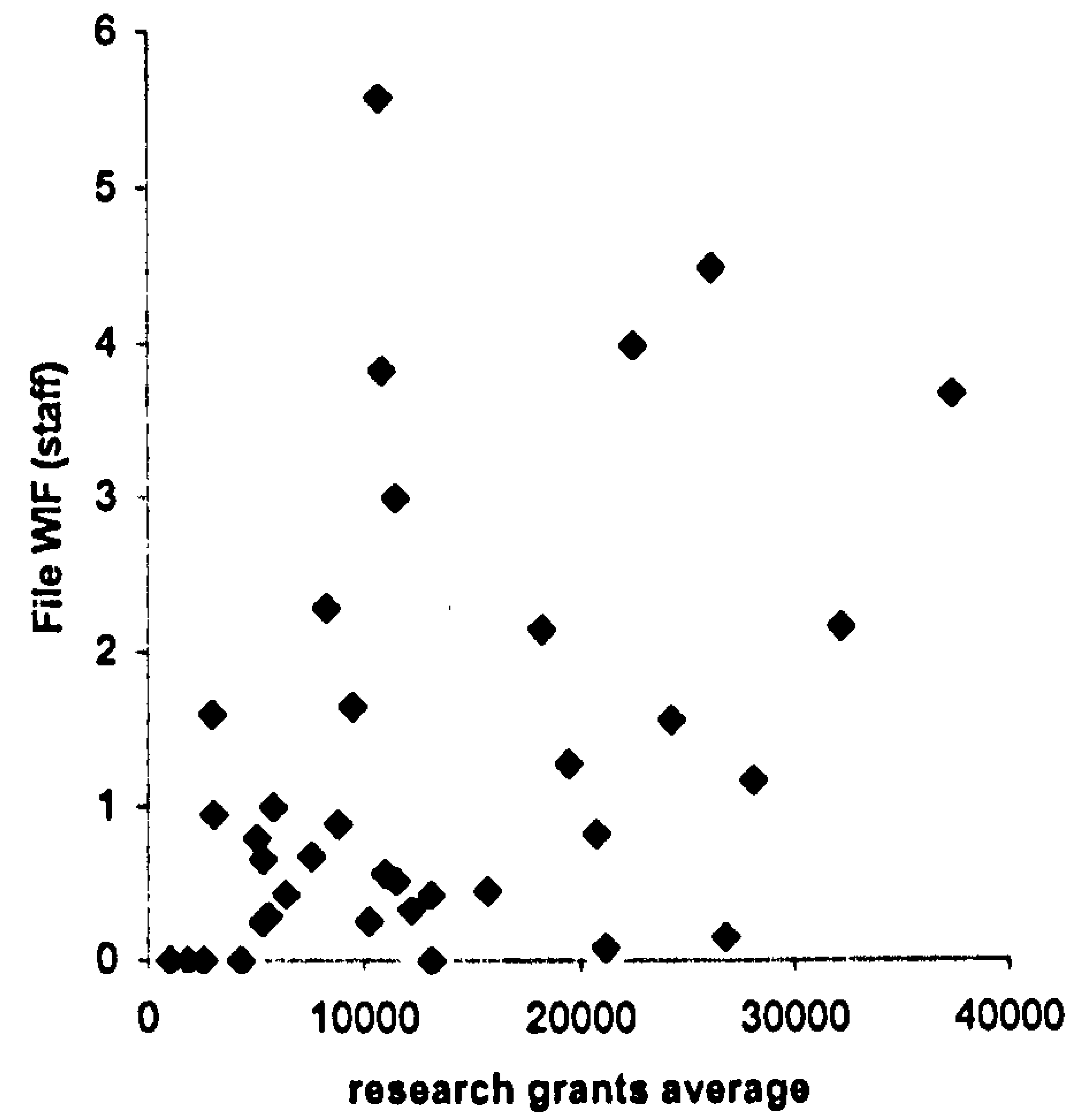


Figure 6.15 File WIFs (staff) against research grant averages for Canadian physics departments

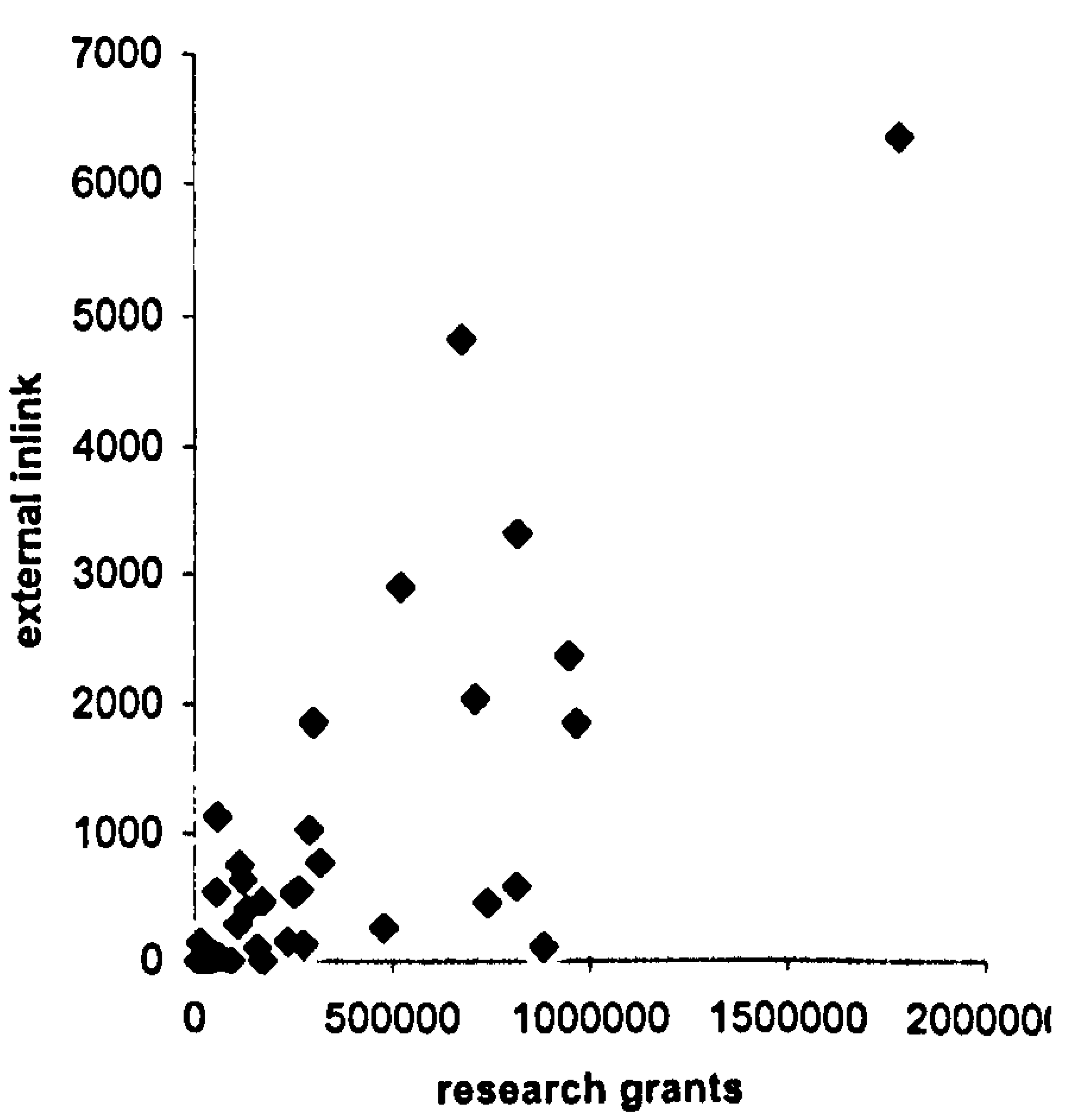


Figure 6.14 External inlinks against research grants for Canadian physics departments

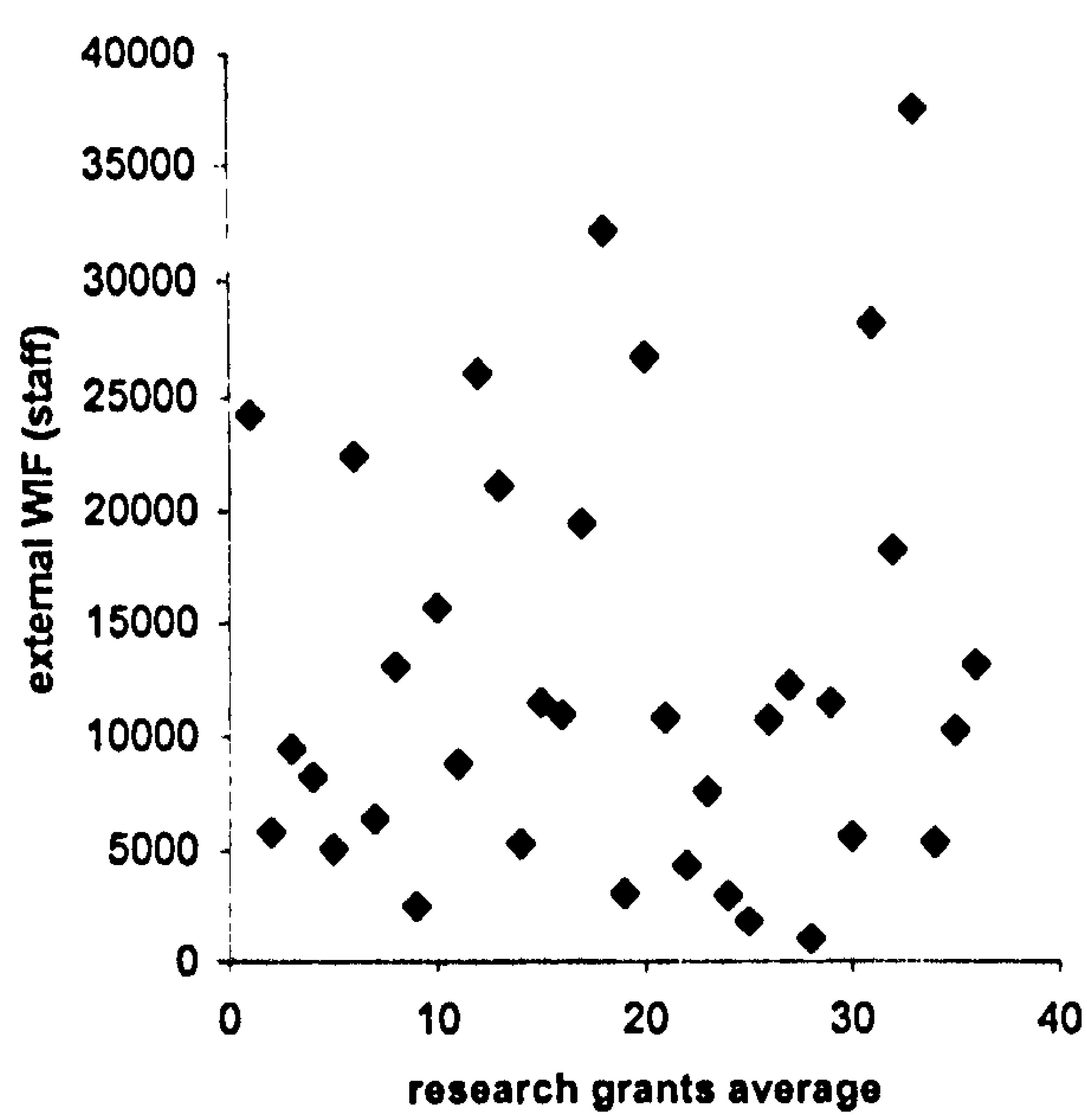


Figure 6.16 External WIFs (staff) against research grant averages for Canadian physics departments

6.1.2.2 Correlation Coefficient Values for the Chemistry Departments

Tables 6.21, 6.22, 6.25 and 6.26 show the Pearson and Spearman correlations between the number of links to or from the chemistry departments with different

ADMs and citations or research grants; and between WIFs or WUFs and citation or research grants averages. Tables 6.23, 6.24, 6.27 and 6.28 list the Pearson and Spearman correlations between the number of inlinks from different domains and citations or research grants; and between WIFs and citation or research averages.

The correlation coefficient values between the inlinks with different ADMs and citations or research grants are nearly all significant at the 1% level in tables 6.21, 6.22, 6.25 and 6.26, with one exception, the Pearson correlation coefficient value between the file inlinks and citations or research grants are not significant at all. The reason is that the chemistry department in the University of Calgary hosts a set of lecture notes created by a lecturer in University of Guelph, and each page of the lecture notes has a link back to its authors' homepage. Altogether the department in the University of Guelph receives 392 links from the University of Calgary for the same reason. Section 8.2.2 discusses this in detail. As the lecture notes are put under one directory, the directory, domain and department document models are not affected by this anomaly.

The correlation coefficient values between the number of inlinks from different domains and citations or research grants are all significant at the 1% level, as shown in tables 6.23, 6.24, 6.27 and 6.28. More significant values are found between different domain WIFs (staff) and citation or research grants averages than those between the WIFs with the numbers of web pages as denominators and citation or research grants averages.

The chemistry department in the University of Guelph appears to be outliers in both figures 6.17 and 6.19. The large number of file inlinks and WIF value for the department dwarf other departments' values.

Table 6.21 Pearson correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=39, n=34 for WIFs and WUFs)

Link measures	File	Directory	Domain	Department
Inlinks	0.230	0.539**	0.664**	0.631**
Outlinks	0.331*	0.297	0.379*	0.314
WIFs (staff)	0.128	0.356*	0.334*	0.219
WUFs (staff)	0.378*	0.171	0.022	-0.091
WIFs	-0.145	-0.084	00097	0.181
WUFs	-0.097	-0.045	-0.042	0.088

Table 6.22 Spearman correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=39, n=34 for WIFs and WUFs)

Link measures	File	Directory	Domain	Department
Inlinks	0.778**	0.779**	0.709**	0.681**
Outlinks	0.374*	0.361*	0.360*	0.334*
WIFs (staff)	0.480**	0.439**	0.323	0.213
WUFs (staff)	-0.088	-0.084	-0.152	-0.188
WIFs	-0.131	-0.108	0.101	0.222
WUFs	0.031	-0.056	0.040	0.079

Table 6.23 Pearson correlations between the link (from AltaVista) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=39, n=37 for WIFs)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.680**	0.641**	0.233
com	0.658**	0.632**	0.240
edu	0.609**	0.591**	0.272
ca	0.703**	0.518**	0.193
org	0.672**	0.635**	0.201
net	0.609**	0.594*	0.289

Table 6.24 Spearman correlations between the link (from AltaVista) and research (citation) measures (* = significant at the 5 % level, ** = significant at the 1% level, n=39, n=37 for WIFs)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.728**	0.461**	0.207
com	0.724**	0.505**	0.219
edu	0.767**	0.650**	0.245
ca	0.631**	0.308	0.195
org	0.767**	0.576**	0.196
net	0.605**	0.411**	0.164

Table 6.25 Pearson correlations between the link (from SocSciBot) and research (research grants) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=40)

Link measures	File	Directory	Domain	Department
Inlinks	0.269	0.540**	0.608**	0.605**
Outlinks	0.221	0.208	0.296	0.247
WIFs (staff)	0.157	0.373*	0.232	0.172
WUFs (staff)	0.327	0.051	-0.139	-0.239
WIFs	0.162	0.178	0.341*	0.522**
WUFs	-0.286	-0.210	-0.039	0.144

Table 6.26 Spearman correlations between the link (from SocSciBot) and research (research grants) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=40)

Link measures	File	Directory	Domain	Department
Inlinks	0.812**	0.808**	0.702**	0.691**
Outlinks	0.388*	0.373*	0.372*	0.352*
WIFs (staff)	0.582**	0.484**	0.261	0.200
WUFs (staff)	-0.180	-0.193	-0.248	-0.285
WIFs	0.148	0.347*	0.501**	0.600**
WUFs	0.061	-0.117	0.026	0.115

Table 6.27 Pearson correlations between the link (from AltaVista) and research (research grants) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=40, when the numbers of web pages are used as denominators n=38)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.601**	0.438**	0.249
com	0.640**	0.444**	0.288
edu	0.521**	0.407**	0.369*
ca	0.628**	0.353*	0.145
org	0.608**	0.453**	0.308
net	0.501**	0.394*	0.262

Table 6.28 Spearman correlations between the link (from AltaVista) and research (research grants) measures (* = significant at the 5 % level, ** = significant at the 1% level, n=40, when the numbers of web pages are used as denominators n=38)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.758**	0.431**	0.124
com	0.753**	0.466**	0.245
edu	0.789**	0.622**	0.475**
ca	0.681**	0.353*	0.037
org	0.804**	0.623**	0.491**
net	0.614**	0.334*	0.094

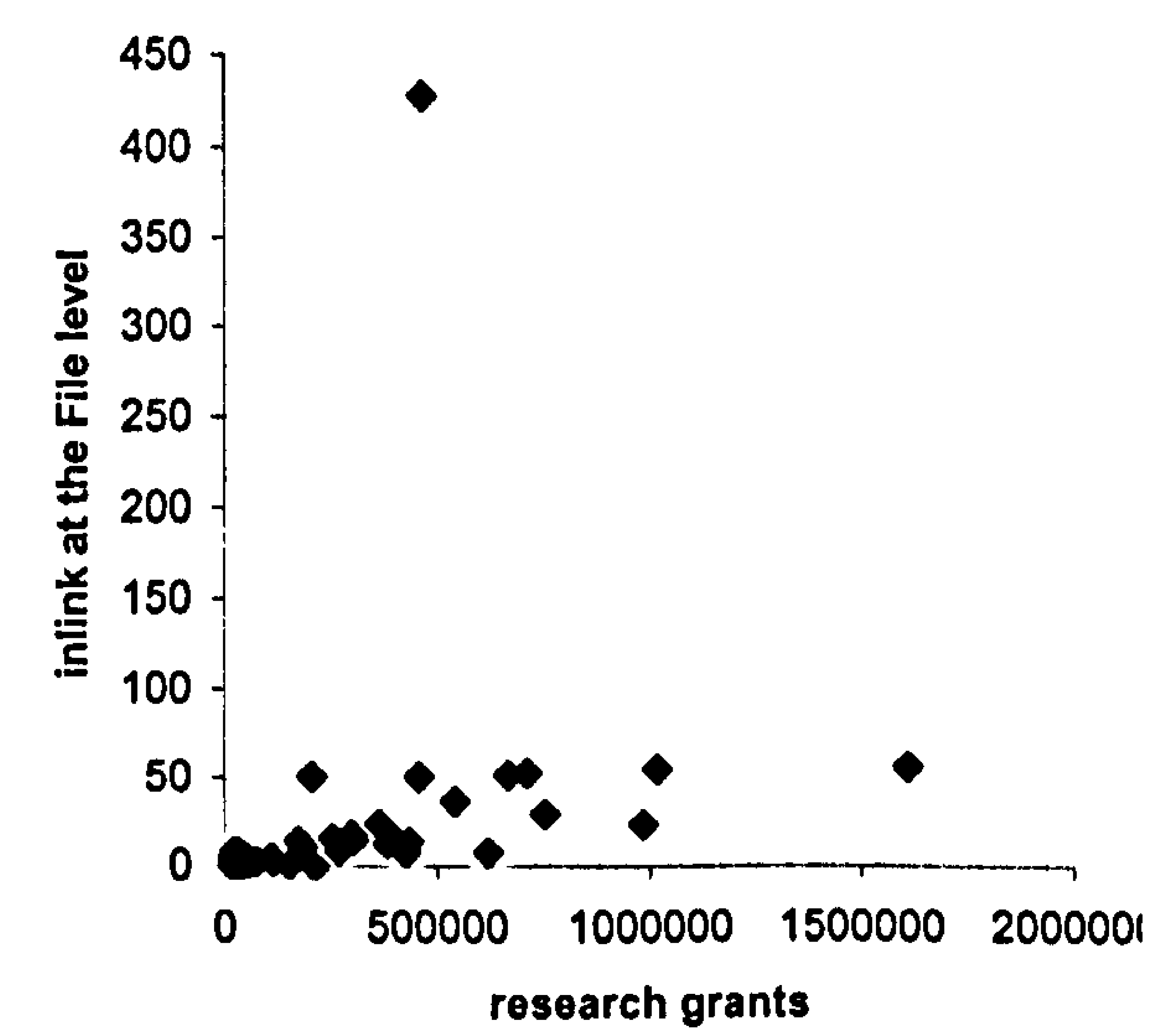


Figure 6.17 File inlinks against research grants for Canadian chemistry departments

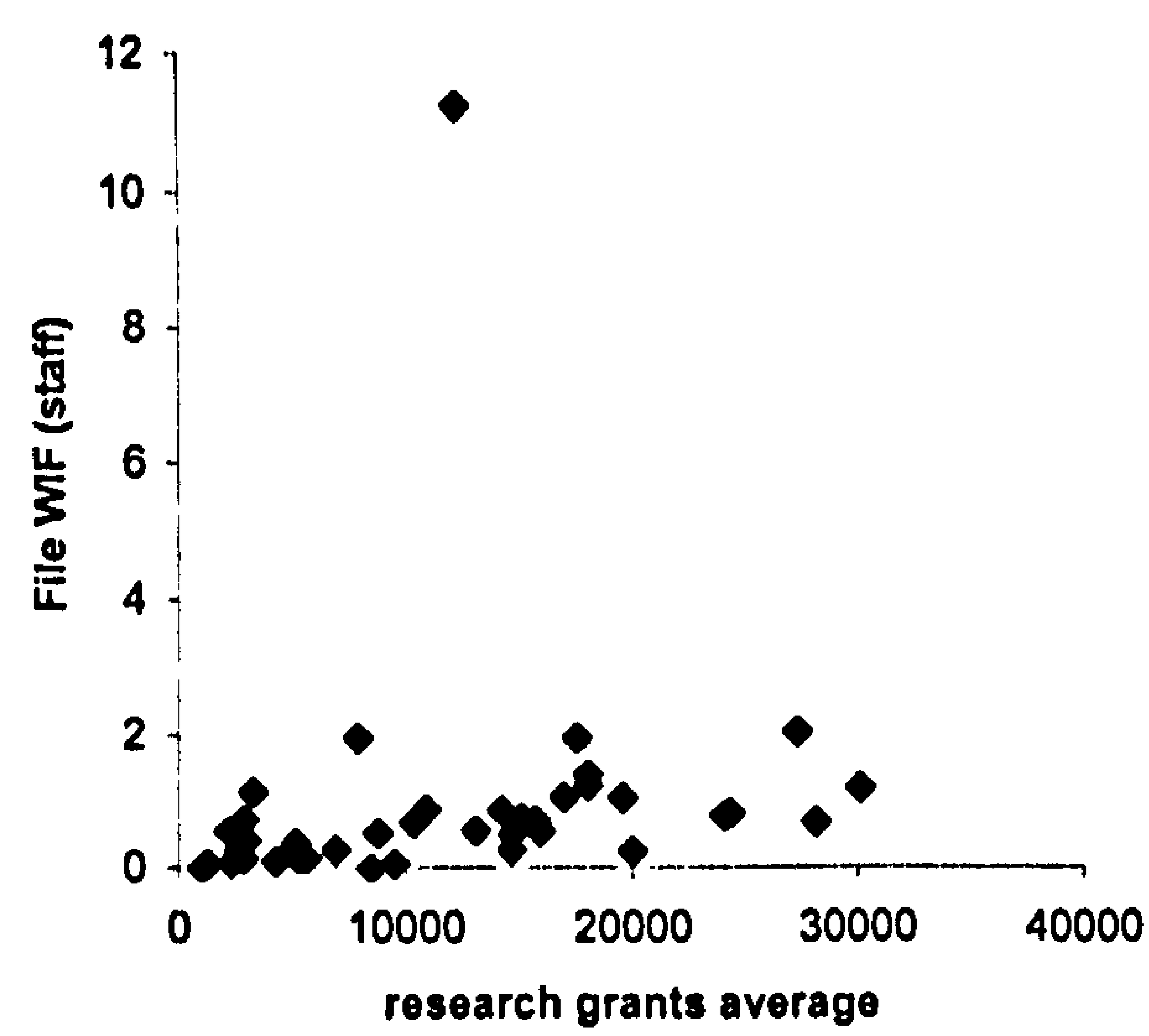
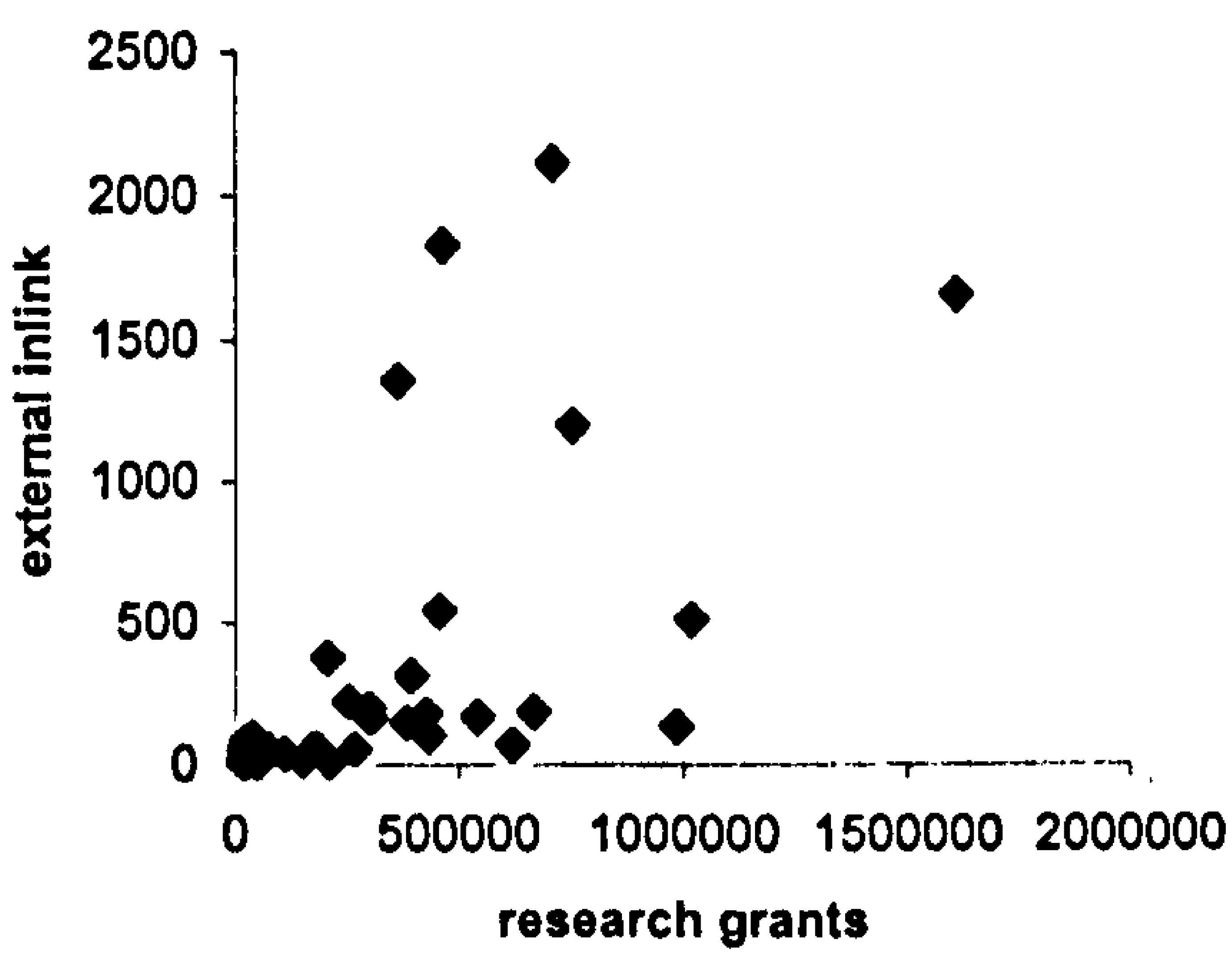


Figure 6.19 File WIFs (staff) against research grant averages for Canadian chemistry departments



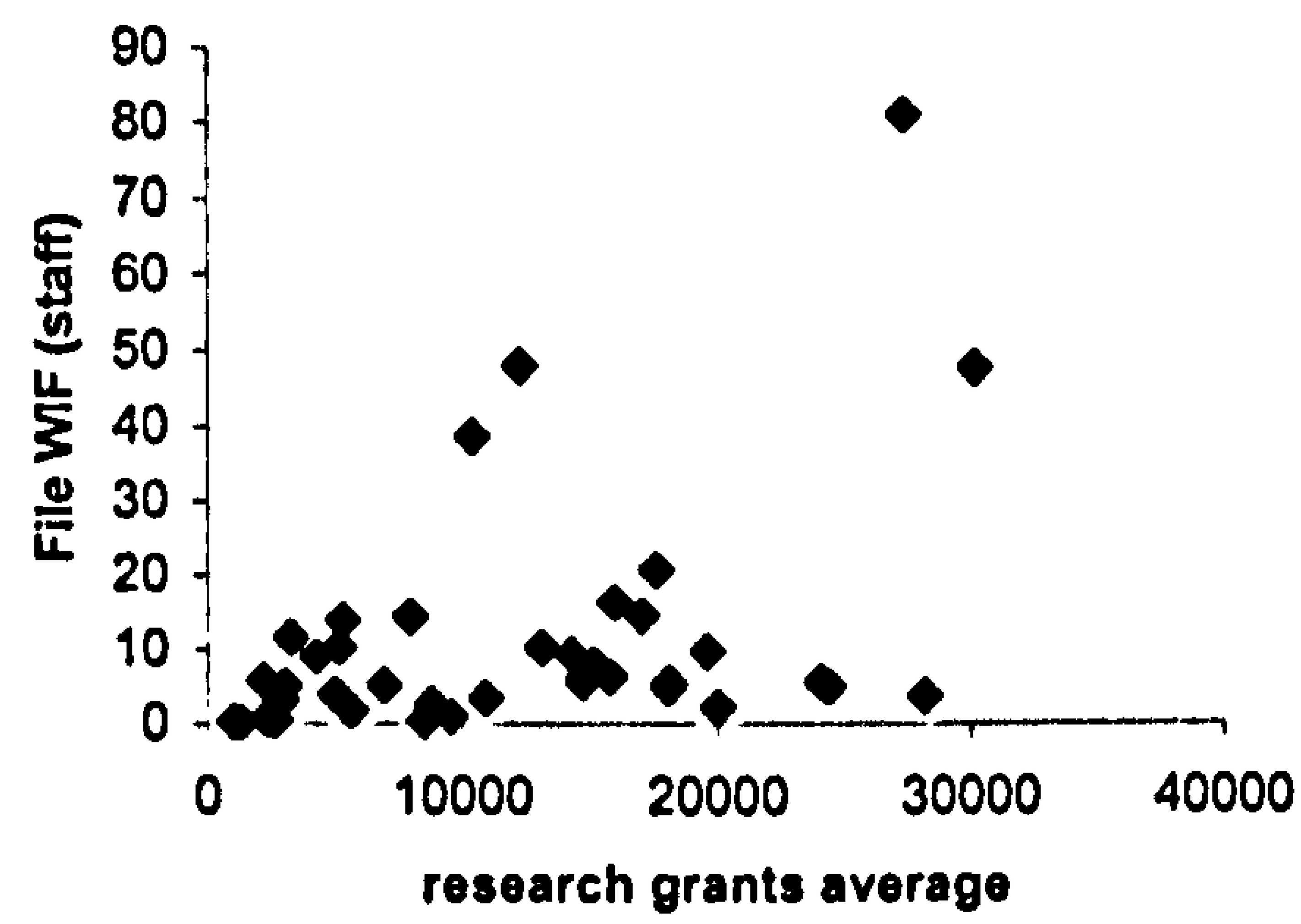


Figure 6.18 External inlinks against research grants for Canadian chemistry departments

Figure 6.20 External WIFs (staff) against research grant averages for Canadian chemistry departments

6.1.2.3 Correlation Coefficient Values for the Biology Departments

Tables 6.29, 6.30, 6.33 and 6.34 list the Pearson and Spearman correlations between the number of links to or from the biology departments with different ADMs and citations or research grants; and between WIFs or WUFs and citation or research grants averages. Tables 6.31, 6.32, 6.35 and 6.36 show the Pearson and Spearman correlations between the number of inlinks from different domains and citations or research grants; and between WIFs and citation or research grants averages.

The correlation coefficient values between links and citations or research grants are nearly all significant at the 1% level, as shown in tables 6.29 to 6.36. Relatively the correlation coefficient values between WIFs or WUFs and citation or research grants averages are less significant. Linear trends are more apparent in figures 6.21 and 6.22 than in figures 6.23 and 6.24.

Table 6.29 Pearson correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=40, n=35 for WIFs and WUFs)

Link measures	File	Directory	Domain	Department
Inlinks	0.482**	0.488**	0.563**	0.608**
Outlinks	0.542**	0.528**	0.570**	0.583**
WIFs (staff)	0.158	0.143	0.133	0.119
WUFs (staff)	0.120	0.176	0.108	0.091
WIFs	-0.153	0.025	0.122	0.489**
WUFs	-0.095	-0.130	0.049	0.511**

Table 6.30 Spearman correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=40, n=35 for WIFs and WUFs)

Link measures	File	Directory	Domain	Department
Inlinks	0.627**	0.627**	0.668**	0.670**
Outlinks	0.549**	0.552**	0.543**	0.543**
WIFs (staff)	0.201	0.200	0.126	0.134
WUFs (staff)	0.114	0.046	-0.097	-0.124
WIFs	0.182	0.213	0.173	0.496**
WUFs	0.094	-0.067	0.010	0.275

Table 6.31 Pearson correlations between the link (from AltaVista) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=40, n=38 for WIFs)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.412**	0.371*	0.286
com	0.376*	0.301	0.304
edu	0.413**	0.359*	-0.017
ca	0.420**	0.326*	0.269
org	0.473**	0.381*	0.489**
net	0.423**	0.462**	0.324*

Table 6.32 Spearman correlations between the link (from AltaVista) and research (citation) measures (* = significant at the 5 % level, ** = significant at the 1% level, n=40, n=38 for WIFs)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.720**	0.431**	0.341*
com	0.711**	0.406**	0.396*
edu	0.645**	0.262	0.036
ca	0.700**	0.466**	0.476**
org	0.742**	0.506**	0.537**
net	0.680**	0.496**	0.390*

Table 6.33 Pearson correlations between the link (from SocSciBot) and research (research grants) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=40)

Link measures	File	Directory	Domain	Department
Inlinks	0.778**	0.763**	0.793**	0.802**
Outlinks	0.399*	0.419**	0.478**	0.489**
WIFs (staff)	0.306	0.298	0.352	0.351
WUFs (staff)	0.038	0.091	0.135	0.132
WIFs	0.073	0.257	0.163	0.159
WUFs	-0.183	-0.286	-0.192	-0.054

Table 6.34 Spearman correlations between the link (from SocSciBot) and research (research grants) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=40)

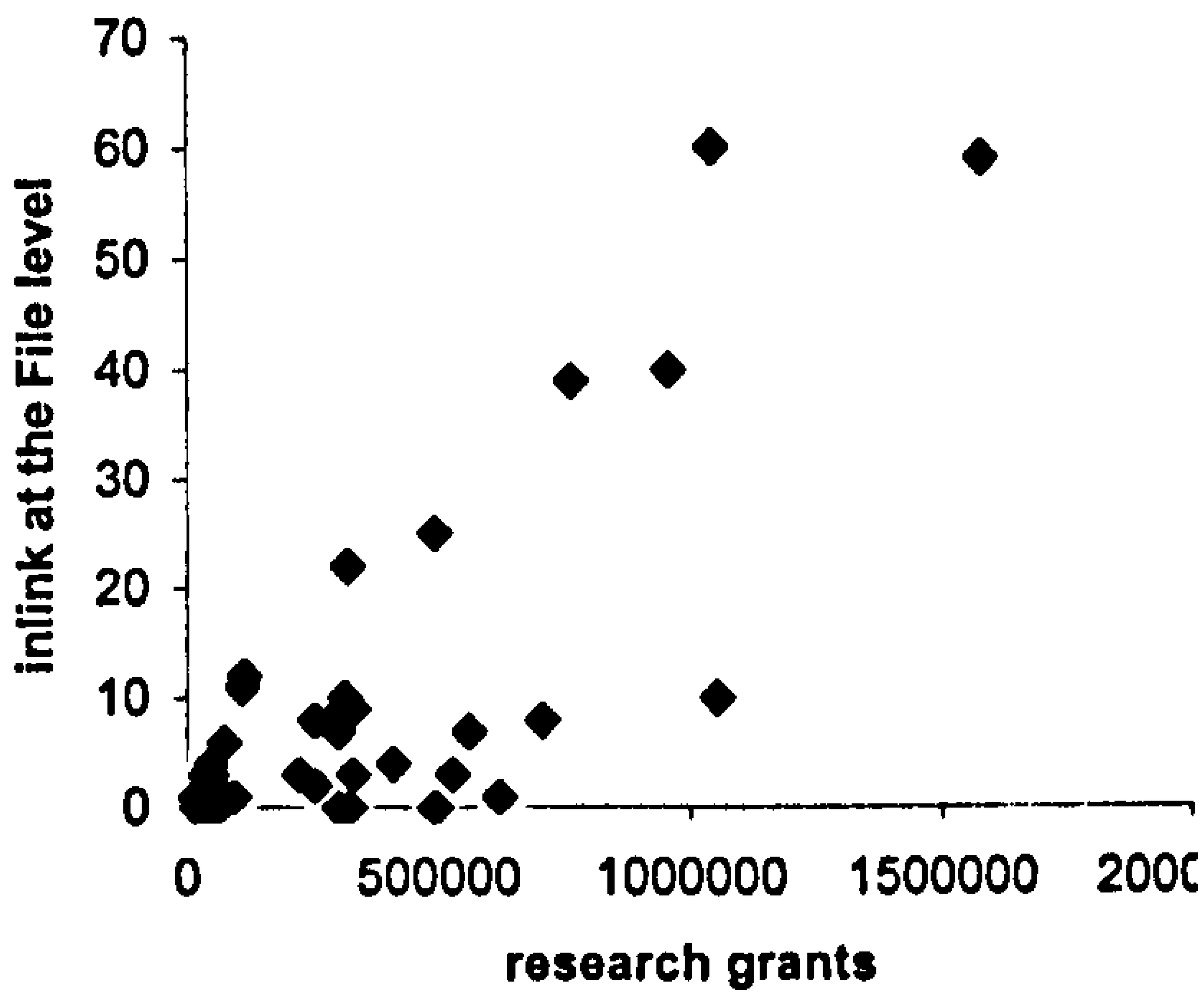
Link measures	File	Directory	Domain	Department
Inlinks	0.607**	0.609**	0.658**	0.661**
Outlinks	0.482**	0.486**	0.489**	0.485**
WIFs (staff)	0.402*	0.408*	0.365*	0.335
WUFs (staff)	0.009	-0.062	-0.062	-0.056
WIFs	0.222	0.310	0.266	0.307
WUFs	-0.041	-0.191	-0.146	-0.090

Table 6.35 Pearson correlations between the link (from AltaVista) and research (research grants) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=40, when the numbers of web pages are used as denominators n=38)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.497**	0.173	0.029
com	0.444**	0.083	-0.030
edu	0.470**	0.167	-0.027
ca	0.492**	0.214	-0.009
org	0.520**	0.112	0.075
net	0.504**	0.239	0.043

Table 6.36 Spearman correlations between the link (from AltaVista) and research (research grants) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=40, when the numbers of web pages are used as denominators n=38)

Source Domain	inlinks and research grants	WIFs (staff) and research grants per staff member	WIFs and research grants per staff member
External	0.610**	0.165	0.142
com	0.590**	0.115	0.088
edu	0.614**	0.163	0.040
ca	0.579**	0.252	0.265
org	0.619**	0.227	0.234
net	0.580**	0.194	0.119



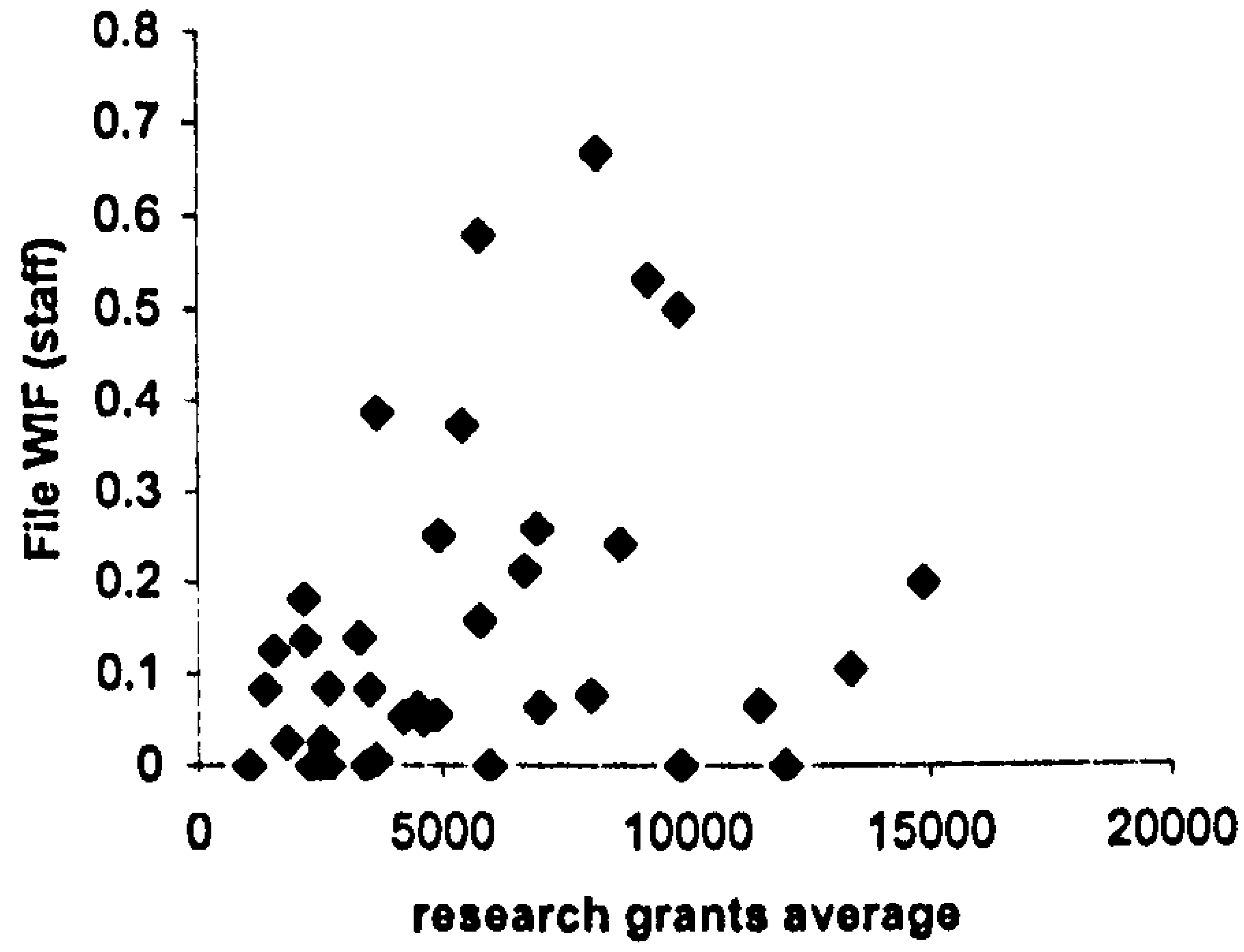
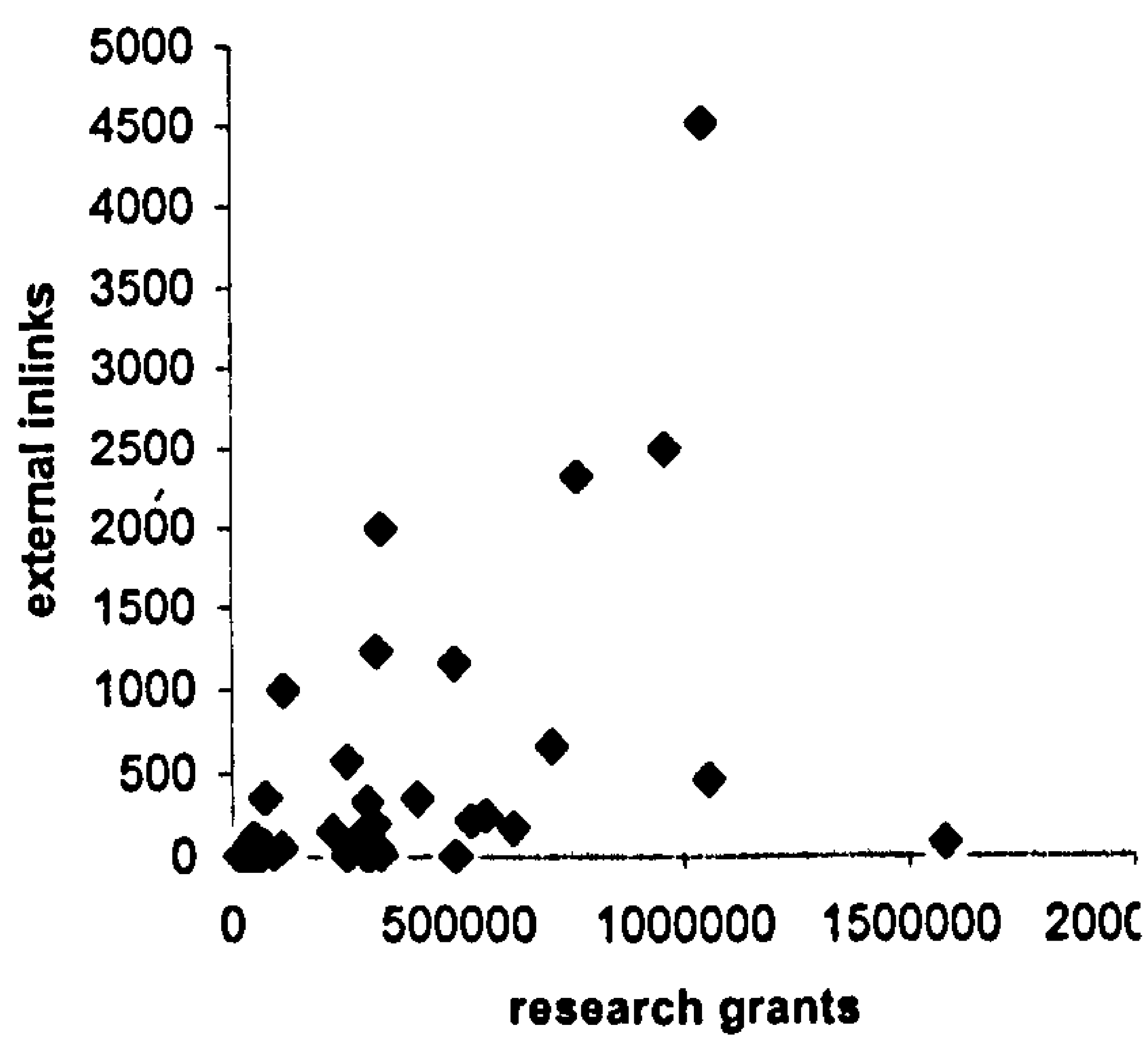


Figure 6.21 File inlinks against research grants for Canadian biology departments

Figure 6.23 File WIFs (staff) against research grant averages for Canadian biology departments



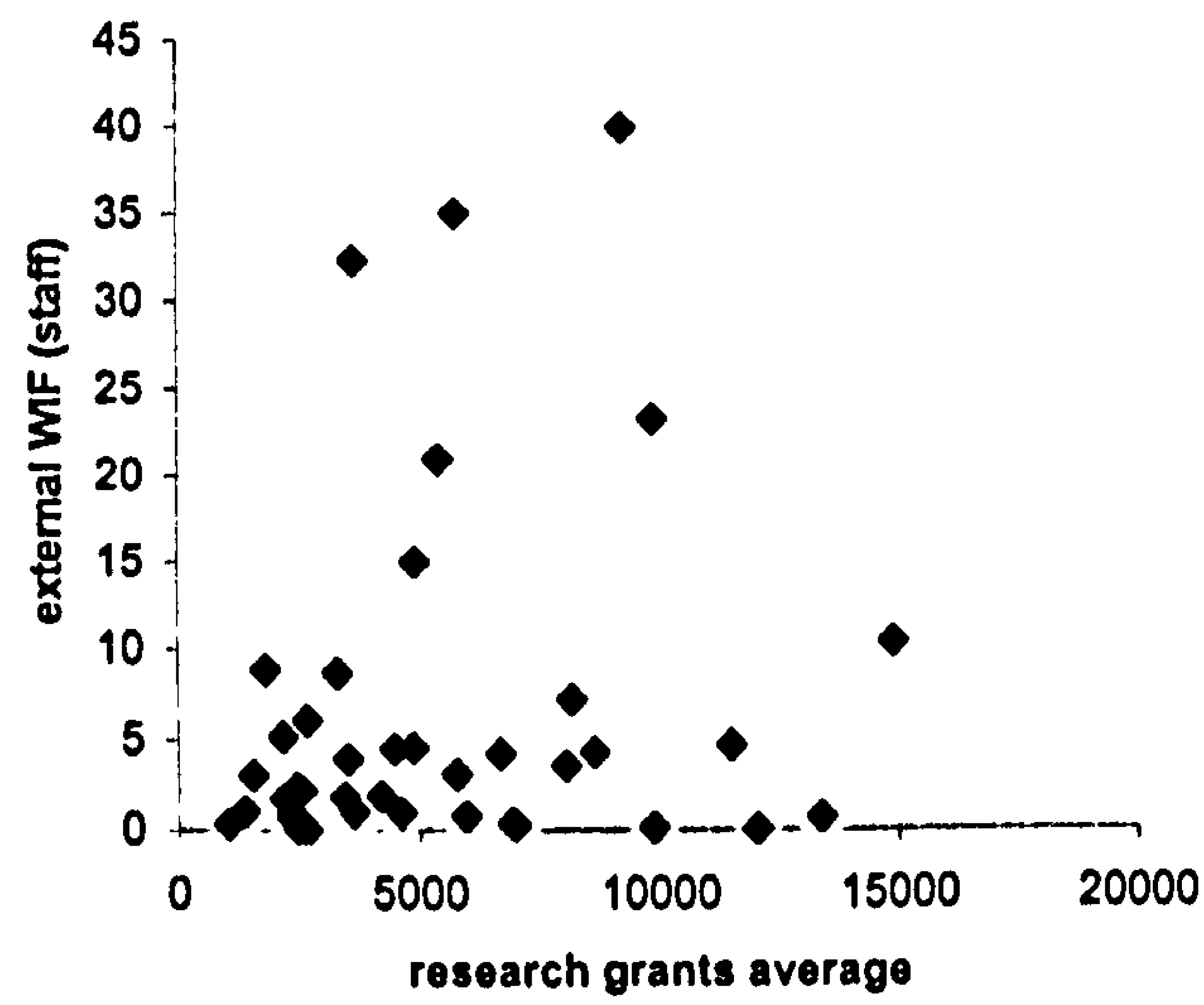


Figure 6.22 External inlinks against research grants for Canadian biology departments

Figure 6.24 *External WIFs (staff) against research grant averages for Canadian biology departments*

6.1.3 Correlation Coefficient Values for UK Departments

6.1.3.1 Correlation Coefficient Values for the Physics Departments

Tables 6.37, 6.38, 6.41 and 6.42 display the Pearson and Spearman correlations between the links to or from the physics departments with different ADMs and citation counts or research productivities (RAE ratings multiply by the staff members submitted); and between WIFs or WUFs and citation or RAE averages. Tables 6.39, 6.40, 6.43 and 6.44 list the Pearson and Spearman correlations between inlinks from different domains and citations or RAE research productivities; and between WIFs and citation or RAE average.

The correlation coefficient values between the inlinks and citations or RAE research productivities are all significant at the 1% level in tables 6.37, 6.38, 6.41 and 6.42. However, the same is not true for the outlinks. There are some significant correlations found between the WIFs (staff) and research averages, although the results are less significant than those between the number of links and research productivities. However, no significant correlations are found between the WIFs (with the numbers of web pages, directories, domains and departments as denominators) and research averages. The Pearson correlation coefficient values are higher than the Spearman ones for the UK physics departments.

The linear trends are more apparent in figures 6.25 and 6.26, which are between the number of inlinks and RAE research productivities than that in figures 6.27 and 6.28, which are between the WIFs (staff) and RAE averages.

Table 6.37 Pearson correlations between the link (from SocSciBot) and research (citation) measures (* = significant at the 5% level, ** = significant at the 1% level, n=42 for WIFs and WUFs, for the rest n=44)

Link measures	File	Directory	Domain	Department
Inlinks	0.435**	0.553**	0.548**	0.479**
Outlinks	0.809**	0.647**	0.647**	0.433**
WIFs (staff)	-0.029	0.088	0.114	0.104
WUFs (staff)	0.142	0.038	0.035	-0.006
WIFs	-0.115	0.231	0.157	0.431**
WUFs	-0.224	-0.220	0.018	0.268

Table 6.38 Spearman correlations between the link (from SocSciBot) and research (citation) measures (* = significant at the 5% level, ** = significant at the 1% level, n=42 for WIFs and WUFs, for the rest n=44)

Link measures	File	Directory	Domain	Department
Inlinks	0.463**	0.463**	0.484**	0.484**
Outlinks	0.359*	0.387**	0.411**	0.392**
WIFs (staff)	0.156	0.217	0.179	0.184
WUFs (staff)	0.213	0.245	0.260	0.231
WIFs	-0.008	0.121	0.054	0.312*
WUFs	-0.258	-0.190	0.011	0.271

Table 6.39 Pearson correlations between the link (from AltaVista) and research (citation) measures (* = significant at the 5% level, ** = significant at the 1% level, n=44)

Source Domain	Inlinks and citations	WIFs(staff) and Citation averages	WIFs and CPP
External	0.844**	0.220	-0.037
com	0.833**	0.175	-0.036
edu	0.865**	0.318*	0.078
uk	0.806**	0.180	-0.279
org	0.848**	0.267	-0.022
net	0.827**	0.231	-0.086
co.uk	0.836**	0.128	-0.239
ac.uk	0.765**	0.167	-0.251

Table 6.40 Spearman correlations between the link (from AltaVista) and research (citation) measures (* = significant at the 5% level, ** = significant at the 1% level, n=44)

Source Domain	Inlinks and citations	WIFs(staff) and Citation averages	WIFs and CPP
External	0.389**	0.169	-0.167
com	0.358*	0.162	-0.120
edu	0.406**	0.243	0.011
uk	0.320*	0.075	-0.364*
org	0.349*	0.143	-0.083
net	0.271	0.081	-0.100
co.uk	0.310*	0.078	-0.315*
ac.uk	0.320*	0.113	-0.335*

Table 6.41 Pearson correlations between the link (from SocSciBot) and research (RAE) measures (* = significant at the 5 % level, ** = significant at the 1% level, n=44)

Link measures	File	Directory	Domain	Department
Inlinks	0.702**	0.745**	0.743**	0.523**
Outlinks	0.638**	0.640**	0.615**	0.404**
WIFs (staff)	0.112	0.222	0.172	0.054
WUFs (staff)	0.129	0.045	-0.070	-0.161
WIFs	-0.243	-0.013	0.095	0.345*
WUFs	-0.317*	-0.190	0.003	0.164

Table 6.42 Spearman correlations between the link (from SocSciBot) and research (RAE) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=44)

Link measures	File	Directory	Domain	Department
Inlinks	0.398**	0.401**	0.404**	0.391**
Outlinks	0.247	0.258	0.251	0.225
WIFs (staff)	0.095	0.140	0.133	0.034
WUFs (staff)	0.122	0.109	0.096	0.019
WIFs	-0.116	-0.025	0.117	0.270
WUFs	-0.101	-0.023	0.118	0.219

Table 6.43 Pearson correlations between the link (from AltaVista) and research (RAE) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=44)

Source domain	inlinks and research productivities	WIFs and RAE averages	WIFs and RAE averages
External	0.647**	0.227	-0.016
com	0.670**	0.271	0.074
edu	0.639**	0.255	0.033
uk	0.710**	0.187	-0.119
org	0.677**	0.281	0.034
net	0.655**	0.252	-0.008
co.uk	0.626**	0.215	-0.088
ac.uk	0.734**	0.157	-0.115

Table 6.44 Spearman correlations between the link (from AltaVista) and research (RAE) measures (* = significant at the 5 % level, ** = significant at the 1% level, n=44)

Source domain	inlinks and research productivities	WIFs (staff) and RAE averages	WIFs and RAE averages
External	0.326*	0.122	-0.115
com	0.303*	0.106	0.005
edu	0.310*	0.196	0.073
uk	0.274	0.067	-0.243
org	0.291	0.154	0.008
net	0.228	0.078	-0.158
co.uk	0.279	0.109	-0.238
ac.uk	0.267	0.075	-0.191

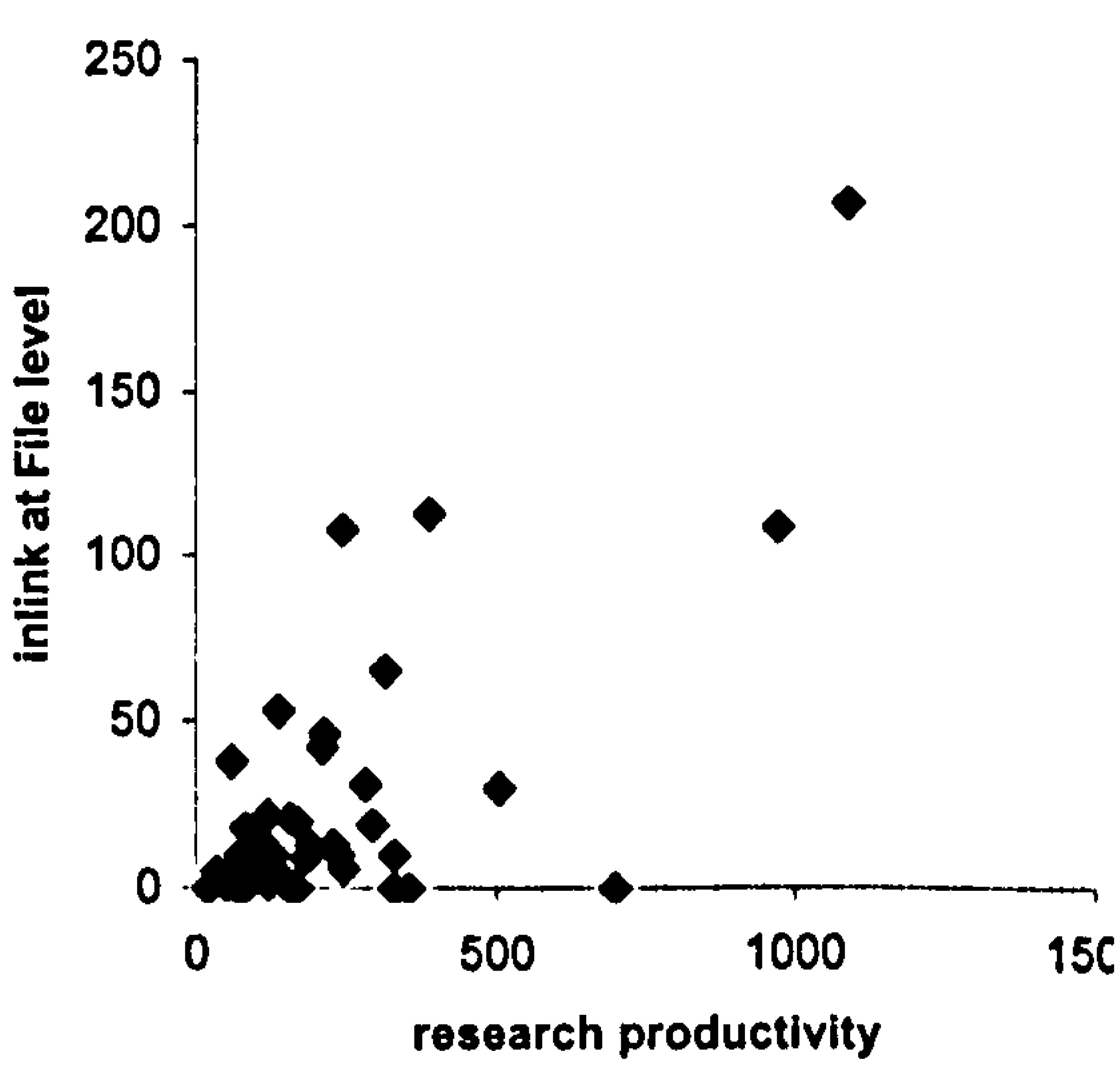


Figure 6.25 File inlinks against research productivities for UK physics departments

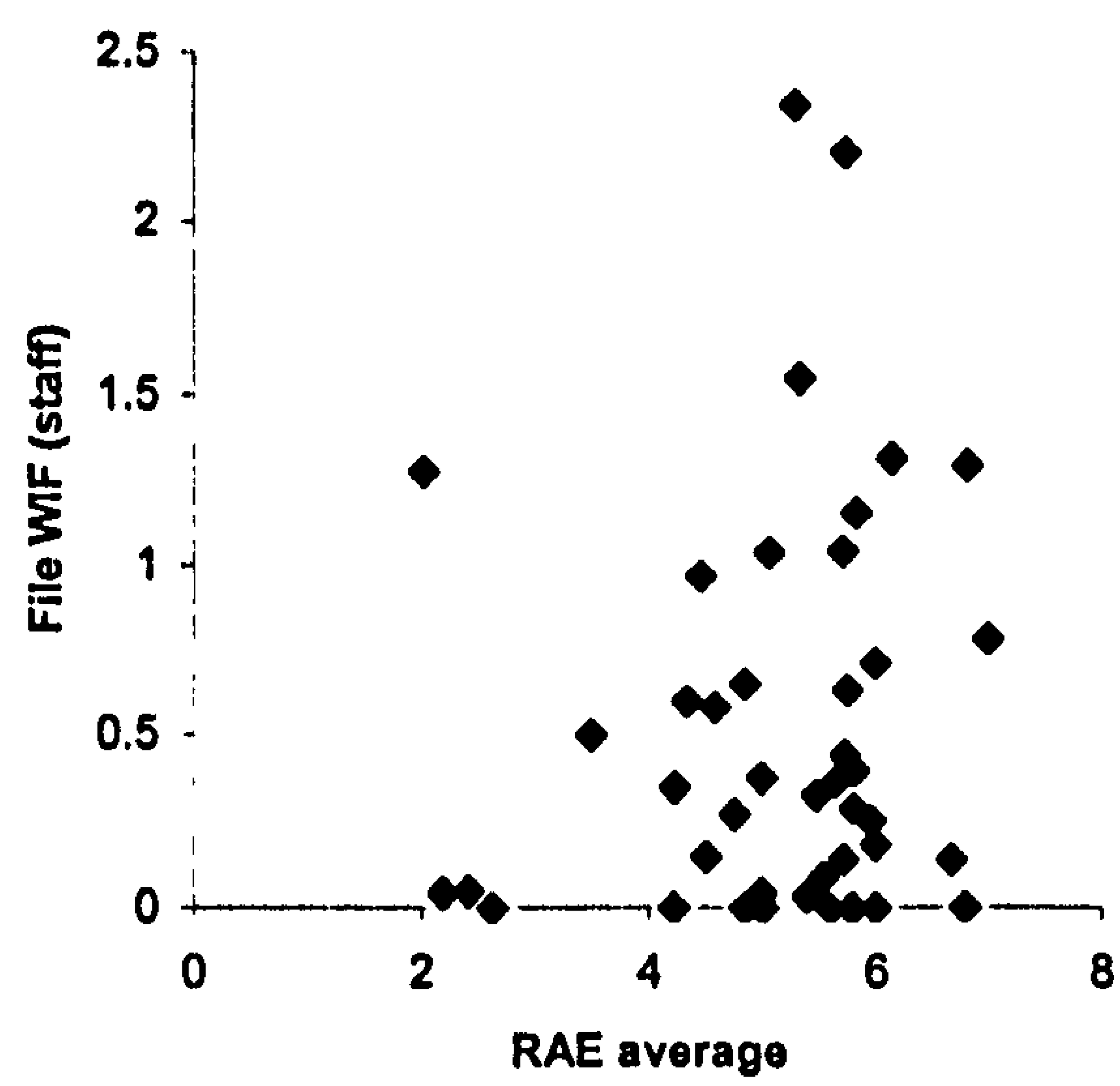
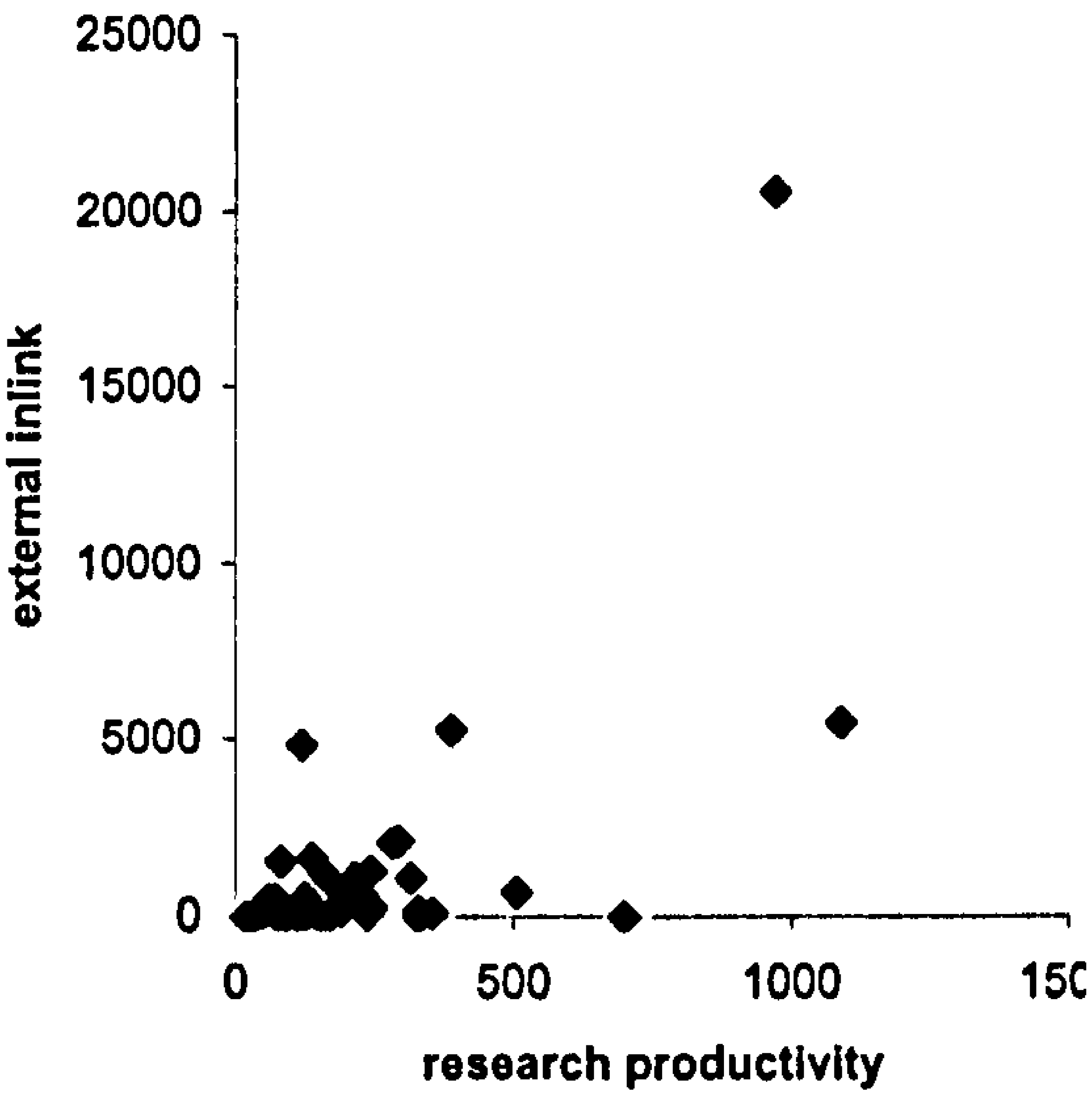


Figure 6.27 File WIFs (staff) against RAE averages for UK physics departments



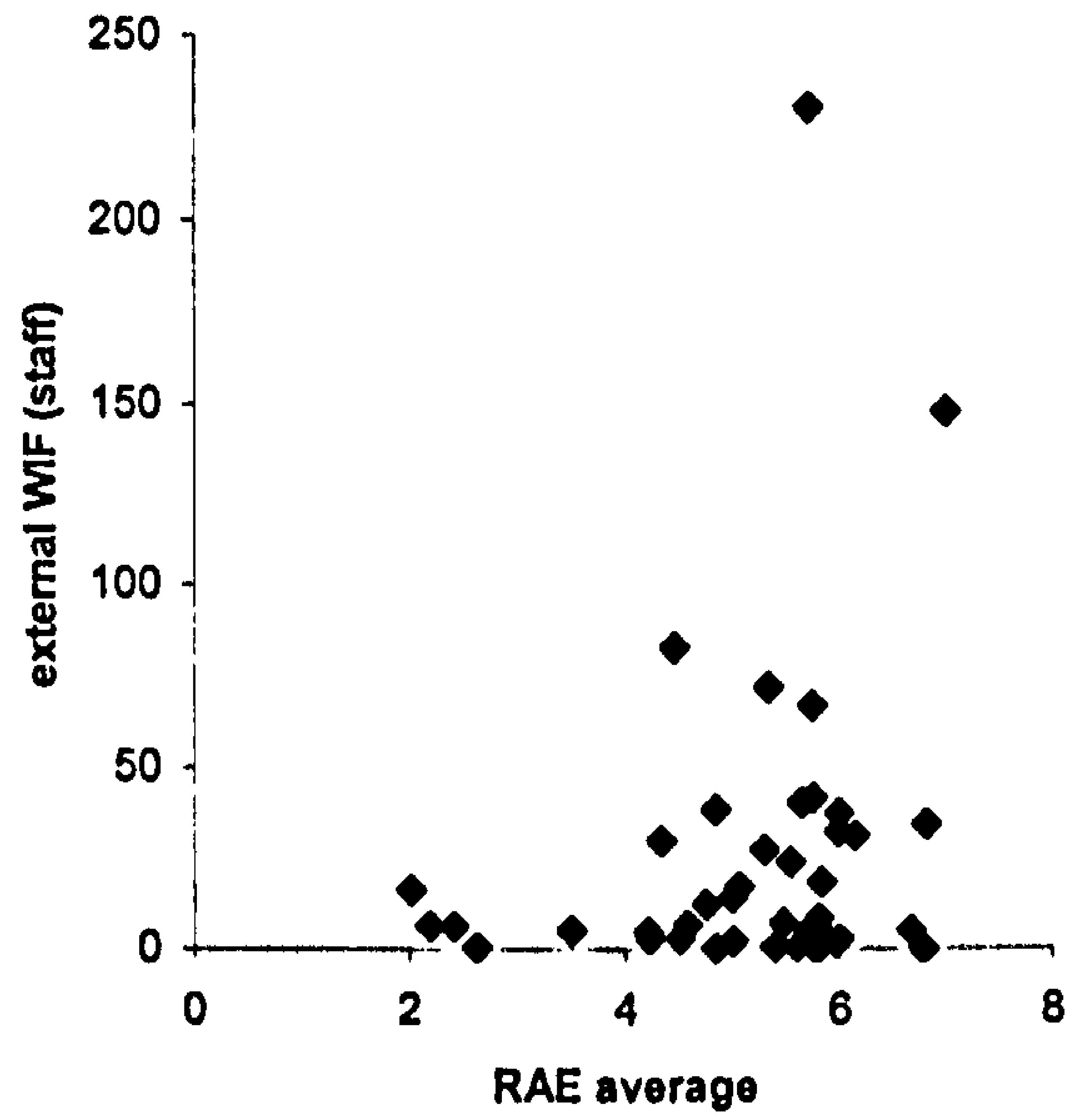


Figure 6.26 External inlinks against research productivities for UK physics departments

Figure 6.28 External WIFs (staff) against RAE averages for UK physics departments

6.1.3.2 Correlation Coefficient Values for the Chemistry Departments

Tables 6.45, 6.46, 6.49 and 6.50 illustrate the Pearson and Spearman correlations between the number of links to or from the chemistry departments with different ADMs and citations or RAE research productivities; and between WIFs or WUFs and citation or RAE averages. Tables 6.47, 6.48, 6.51 and 6.52 display the Pearson and Spearman correlations between the inlinks from different domains and citations or RAE research productivities; and between WIFs and citation or RAE averages.

The correlation coefficient values between the number of links and citations or RAE research productivities are nearly all significantly at the 1% level in tables 6.45 to 6.52. Fewer significant values are found between the WIFs or WUFs and citation or RAE averages.

The linear trends in figures 6.29 and 6.30, between the number of inlinks and RAE research productivities are more apparent than those in figures 6.31 and 6.32, between the WIFs (staff) and RAE averages.

Table 6.45 Pearson correlations between the link (from SocSciBot) and research (citation) measures (* = significant at the 5% level, ** = significant at the 1% level, n=38 for WIFs and WUFs, for the rest n=41)

Link measures	File	Directory	Domain	Department
Inlinks	0.626**	0.441**	0.577**	0.536**
Outlinks	0.576**	0.619**	0.574**	0.476**
WIFs (staff)	0.196	0.154	0.108	0.016
WUFs (staff)	0.235	0.220	0.156	0.045
WIFs	-0.165	0.174	0.270	0.530**
WUFs	0.195	0.364*	0.279	0.442**

Table 6.46 Spearman correlations between the link (from SocSciBot) and research (citation) measures (* = significant at the 5% level, ** = significant at the 1% level, n=38 for WIFs and WUFs, for the rest n=41)

Link measures	File	Directory	Domain	Department
Inlinks	0.567**	0.522**	0.553**	0.525**
Outlinks	0.285	0.289	0.332*	0.313*
WIFs (staff)	0.236	0.325*	0.190	0.098
WUFs (staff)	0.077	0.067	0.066	0.023
WIFs	-0.023	0.111	0.256	0.521**
WUFs	0.074	0.299	0.308	0.423**

Table 6.47 Pearson correlations between the link (from AltaVista) and research (citation) measures (* = significant at the 5% level, ** = significant at the 1% level, n=41)

Source Domain	Inlinks and citations	WIFs(staff) and citation averages	WIFs and CPP
External	0.495**	0.008	0.249
com	0.425**	0.063	0.266
edu	0.367*	-0.038	0.241
uk	0.521**	0.050	0.187
org	0.476**	0.013	0.201
net	0.541**	0.051	0.184
co.uk	0.665**	0.211	0.176
ac.uk	0.492**	0.034	0.185

Table 6.48 Spearman correlations between the link (from AltaVista) and research (citation) measures (* = significant at the 5% level, ** = significant at the 1% level, n=41)

Source Domain	Inlinks and citations	WIFs(staff) and citation averages	WIFs and CPP
External	0.383*	0.088	0.329*
com	0.284	0.104	0.359*
edu	0.340*	0.123	0.412**
uk	0.426**	0.143	0.267
org	0.285	0.046	0.126
net	0.251	0.020	0.167
co.uk	0.299	0.110	0.238
ac.uk	0.442**	0.152	0.259

Table 5.49 Pearson correlations between the link (from SocSciBot) and research (RAE) measures (* = significant at the 5 % level, ** = significant at the 1% level, n=41)

Link measures	File	Directory	Domain	Department
Inlinks	0.786**	0.618**	0.742**	0.730**
Outlinks	0.495**	0.602**	0.619**	0.573**
WIFs (staff)	0.399**	0.374*	0.306	0.181
WUFs (staff)	0.188	0.281	0.260	0.122
WIFs	0.118	0.229	0.236	0.615**
WUFs	0.014	0.415**	0.190	0.480**

Table 5.50 Spearman correlations between the link (from SocSciBot) and research (RAE) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=41)

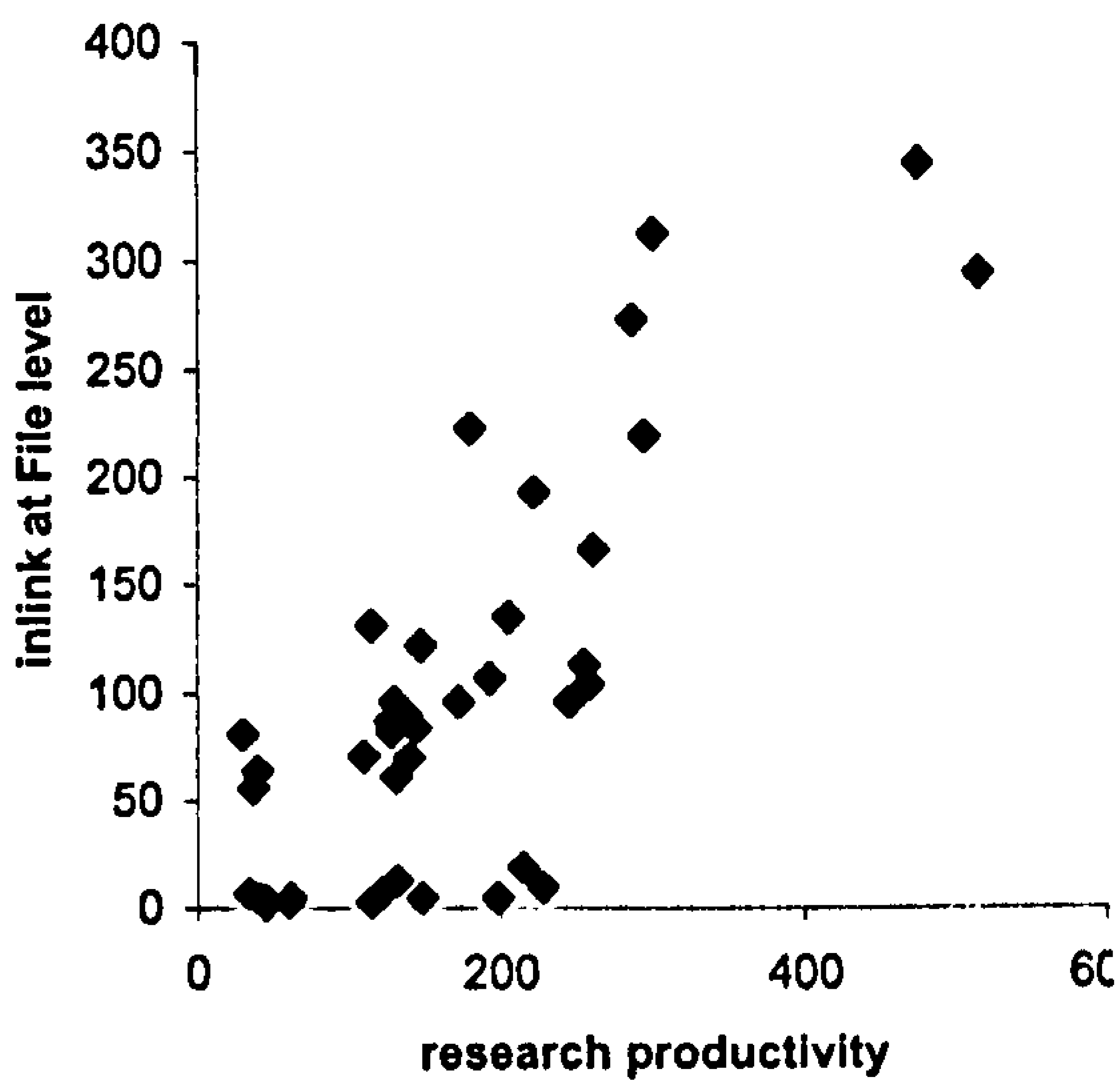
Link measures	File	Directory	Domain	Department
Inlinks	0.708**	0.640**	0.702**	0.682**
Outlinks	0.450**	0.468**	0.502**	0.478**
WIFs (staff)	0.377*	0.419**	0.274	0.182
WUFs (staff)	0.234	0.219	0.168	0.123
WIFs	0.076	0.147	0.265	0.625**
WUFs	0.120	0.378*	0.305	0.462**

Table 5.51 Pearson correlations between the link (from AltaVista) and research (RAE) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=41)

Source Domain	inlinks and research productivities	WIFs (staff) and RAE averages	Between WIFs and RAE averages
External	0.642**	0.188	0.147
com	0.504**	0.148	0.140
edu	0.500**	0.120	0.195
uk	0.675**	0.248	-0.039
org	0.603**	0.196	0.219
net	0.672**	0.279	0.227
co.uk	0.772**	0.420**	0.057
ac.uk	0.648**	0.223	-0.046

Table 5.52 Spearman correlations between the link (from AltaVista) and research (RAE) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=41)

Source Domain	inlinks and research productivities	WIFs (staff) and RAE averages	Between WIFs and RAE averages
External	0.586**	0.318*	0.175
com	0.432**	0.270	0.174
edu	0.541**	0.369*	0.249
uk	0.642**	0.369*	0.159
org	0.500**	0.299	0.191
net	0.485**	0.253	0.293
co.uk	0.488**	0.331*	0.124
ac.uk	0.653**	0.372*	0.166



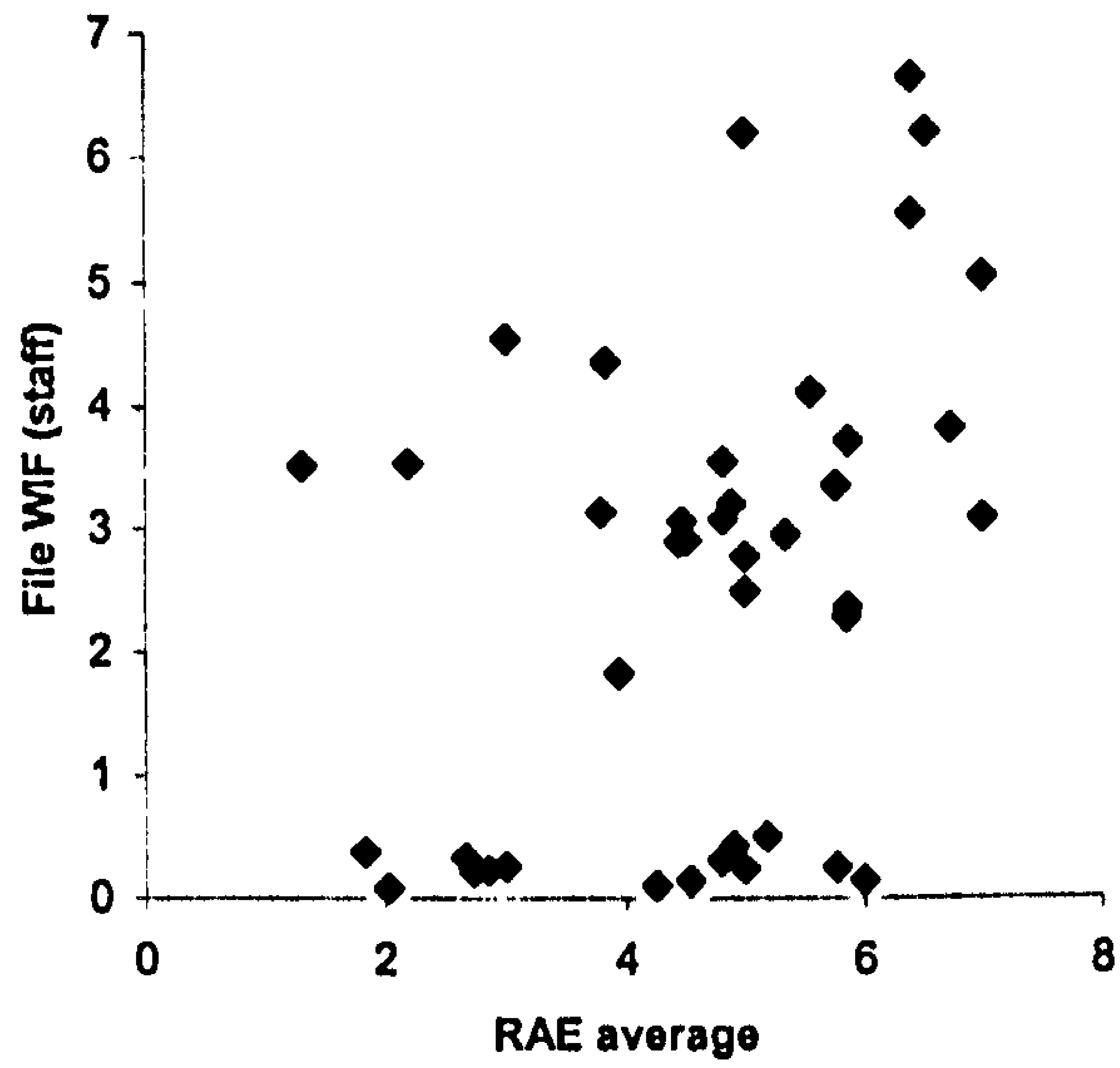
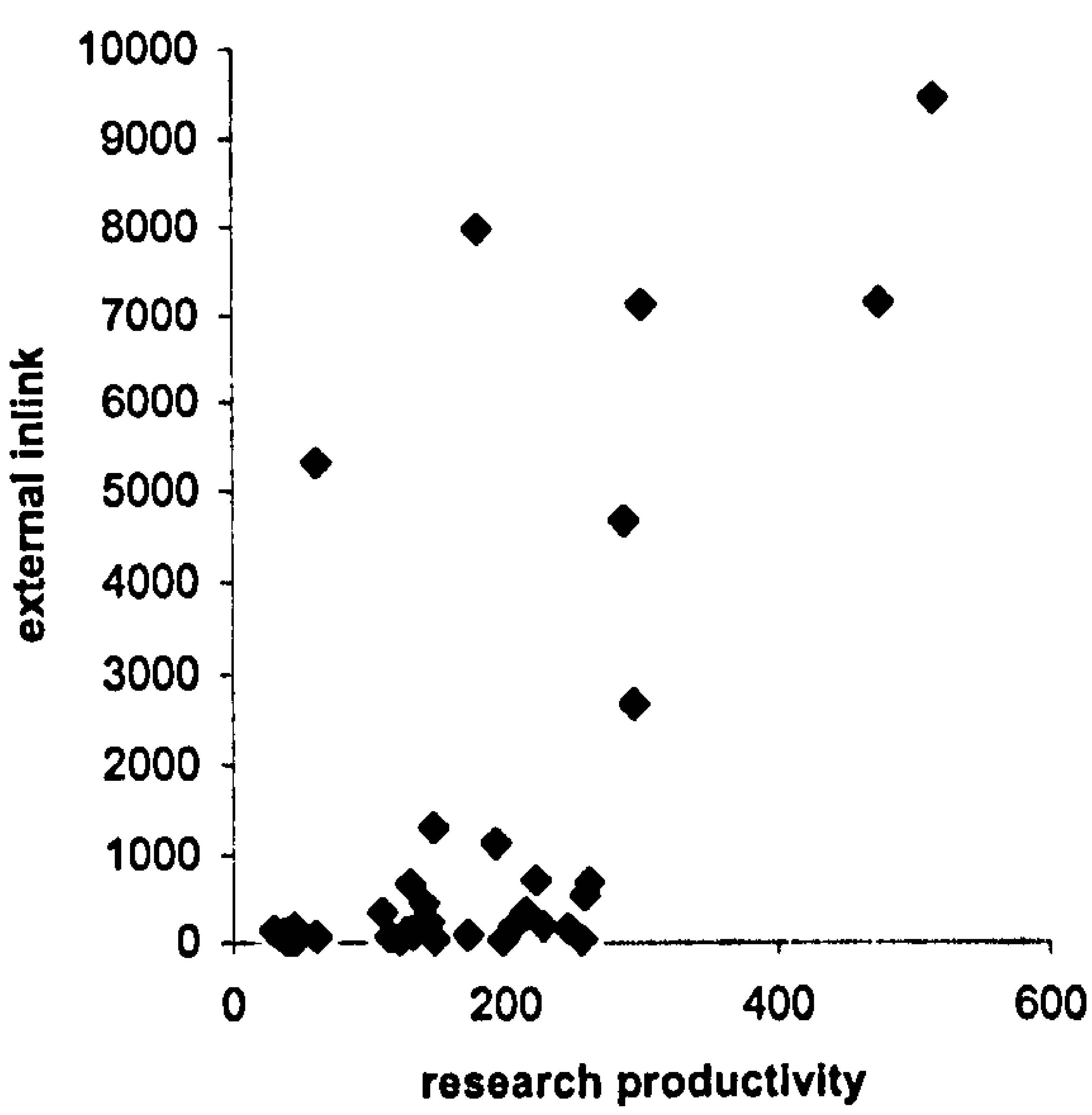


Figure 6.29 File inlinks against research productivities for UK chemistry departments

Figure 6.31 File WIFs (staff) against RAE averages for UK chemistry departments



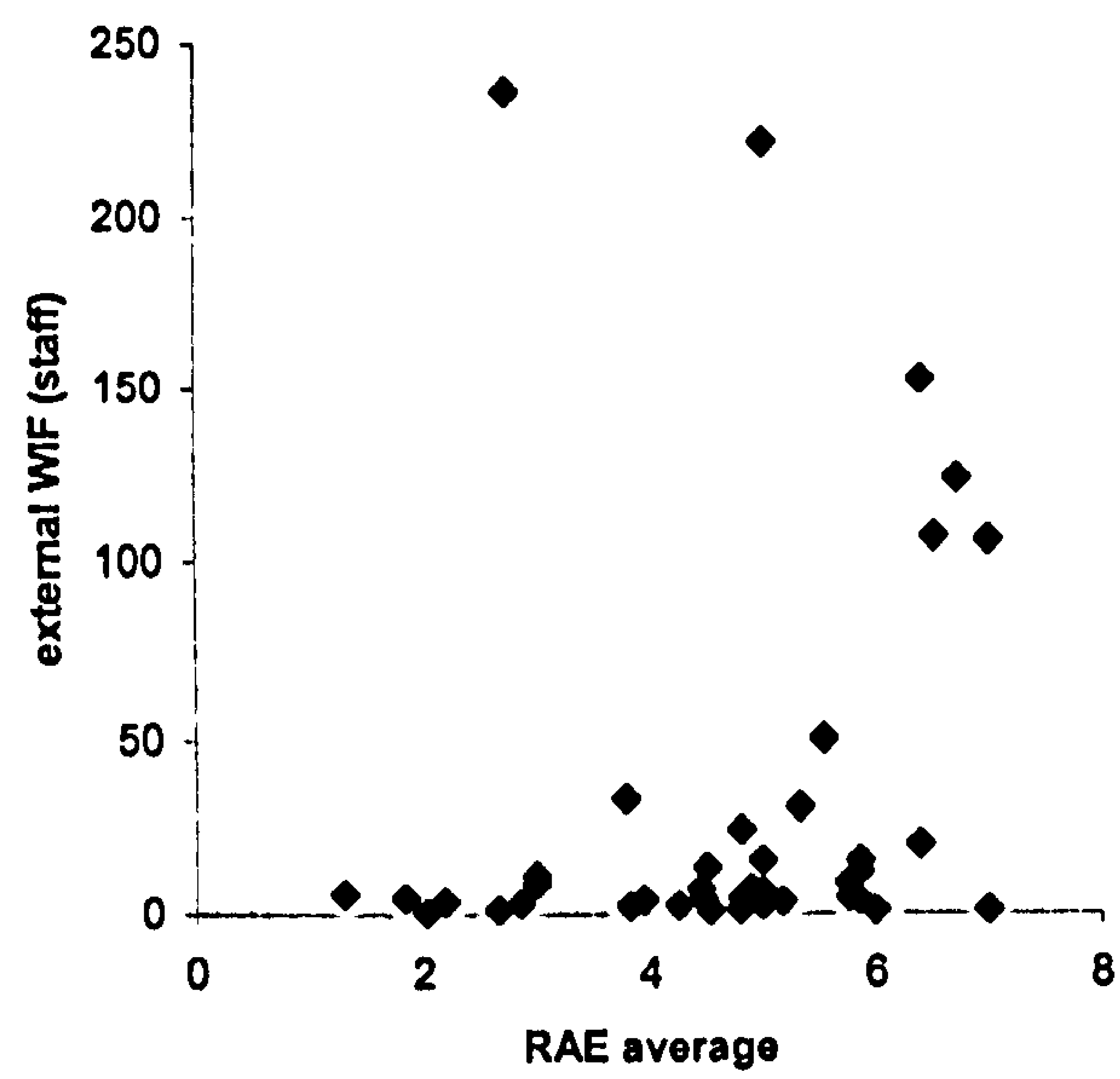


Figure 6.30 External inlinks against the research productivities for UK chemistry departments

***Figure 6.32** External WIFs (staff) against RAE averages for UK chemistry departments*

6.1.3.3 Correlation Coefficient Values for the Biology Departments

Tables 6.53, 6.54, 6.57 and 6.58 display the Pearson and Spearman correlations between the number of links to or from the biology departments with different ADMs and citations or RAE research productivities; and between WIFs or WUFs and citation or RAE averages. Tables 6.55, 6.56, 6.59 and 6.60 illustrate the Pearson and Spearman correlations between the inlinks from different domains and citations or RAE research productivities, and between WIFs and citation or RAE averages.

The Spearman correlations are more significant than the Pearson correlations as shown in tables 6.53 to 6.60. The Pearson correlation coefficient values between the file or directory inlinks and research productivities are not at all significant, as shown in table 6.57. This is caused by the biology department in the University College London, which hosts a lot of databases and attracts an extremely large number of links from its peers in the UK. However, the domain and departments document models successfully remove this effect.

In tables 6.54 and 6.58, the correlation coefficient values are nearly all significant at the 1% level, even the values between the WIFs or WUFs (with number of web pages, directories, domains and departments as denominators) and research averages.

The number of inlinks attracted by the biology department in the University College London dwarfs other departments' values, especially in figures 6.33 and 6.35. It appears to be apparent outliers in all the four figures from 6.33 to 6.36.

Table 6.53 Pearson correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, for WIFs and WUFs n=46, the rest n=53)

Link measures	File	Directory	Domain	Department
Inlinks	0.432**	0.428**	0.624**	0.681**
Outlinks	0.385**	0.381**	0.450**	0.501**
WIFs (staff)	0.309*	0.308*	0.195	0.185
WUFs (staff)	0.089	0.087	0.124	0.165
WIFs	0.332*	0.344*	0.514**	0.587**
WUFs	0.216	0.138	0.325*	0.385*

Table 6.54 Spearman correlations between the link (from SocSciBot) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, for WIFs and WUFs n=46, the rest n=53)

Link measures	File	Directory	Domain	Department
Inlinks	0.673**	0.673**	0.687**	0.673**
Outlinks	0.511**	0.535**	0.591**	0.577**
WIFs (staff)	0.477**	0.454**	0.411**	0.369**
WUFs (staff)	0.303*	0.344*	0.366**	0.337*
WIFs	0.508**	0.492**	0.546**	0.643**
WUFs	0.383**	0.418**	0.432**	0.588**

Table 6.55 Pearson correlations between the link (from AltaVista) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=53)

Source Domain	inlinks and citations	WIFs (staff) and citation averages	WIFs and CPP
External	0.589**	0.258	-0.153
com	0.546**	0.099	-0.181
edu	0.528**	0.262	-0.065
uk	0.708**	0.264	-0.154
org	0.650**	0.184	0.128
net	0.563**	0.248	0.075
co.uk	0.562**	0.147	-0.164
ac.uk	0.684**	0.294*	-0.144

Table 6.56 Spearman correlations between the link (from AltaVista) and research (citation) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=53)

Source Domain	inlinks and citations	WIFs (staff) and citation averages	WIFs and CPP
External	0.661**	0.369**	0.030
com	0.568**	0.253	0.053
edu	0.644**	0.380**	0.300*
uk	0.688**	0.381**	0.053
org	0.632**	0.361**	0.322*
net	0.569**	0.303*	0.285*
co.uk	0.620**	0.292*	0.106
ac.uk	0.709**	0.444**	0.108

Table 6.57 Pearson correlations between the link (from SocSciBot) and research (RAE) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=53)

Link measures	File	Directory	Domain	Department
Inlinks	0.196	0.191	0.556**	0.651**
Outlinks	0.471**	0.470**	0.512**	0.595**
WIFs (staff)	0.132	0.130	0.279*	0.360**
WUFs (staff)	0.163	0.166	0.198	0.301*
WIFs	0.123	0.157	0.488**	0.500**
WUFs	0.303*	0.196	0.444**	0.438**

Table 6.58 Spearman correlations between the link (from SocSciBot) and research (RAE) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=53)

Link measures	File	Directory	Domain	Department
Inlinks	0.790**	0.789**	0.791**	0.779**
Outlinks	0.663**	0.657**	0.701**	0.690**
WIFs (staff)	0.602**	0.598**	0.557**	0.496**
WUFs (staff)	0.473**	0.465**	0.474**	0.452**
WIFs	0.482**	0.512**	0.586**	0.630**
WUFs	0.506**	0.421**	0.469**	0.570**

Table 6.59 Pearson correlations between the link (from AltaVista) and research (RAE) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=53)

Source Domain	inlinks and research productivities	WIFs (staff) and RAE averages	WIFs and RAE averages
External	0.442**	0.208	-0.308*
com	0.587**	0.214	-0.269
edu	0.355**	0.185	-0.193
uk	0.630**	0.276*	-0.327*
org	0.669**	0.254	0.202
net	0.432**	0.212	0.132
co.uk	0.621**	0.238	-0.305*
ac.uk	0.562**	0.284*	-0.307*

Table 6.60 Spearman correlations between the link (from AltaVista) and research (RAE) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=53)

Source Domain	inlinks and research productivities	WIFs (staff) and RAE averages	WIFs and RAE averages
External	0.755**	0.493**	0.069
com	0.654**	0.364**	0.063
edu	0.753**	0.521**	0.329*
uk	0.764**	0.452**	-0.011
org	0.734**	0.483**	0.343*
net	0.692**	0.453**	0.307*
co.uk	0.708**	0.390**	0.039
ac.uk	0.767**	0.529**	0.075

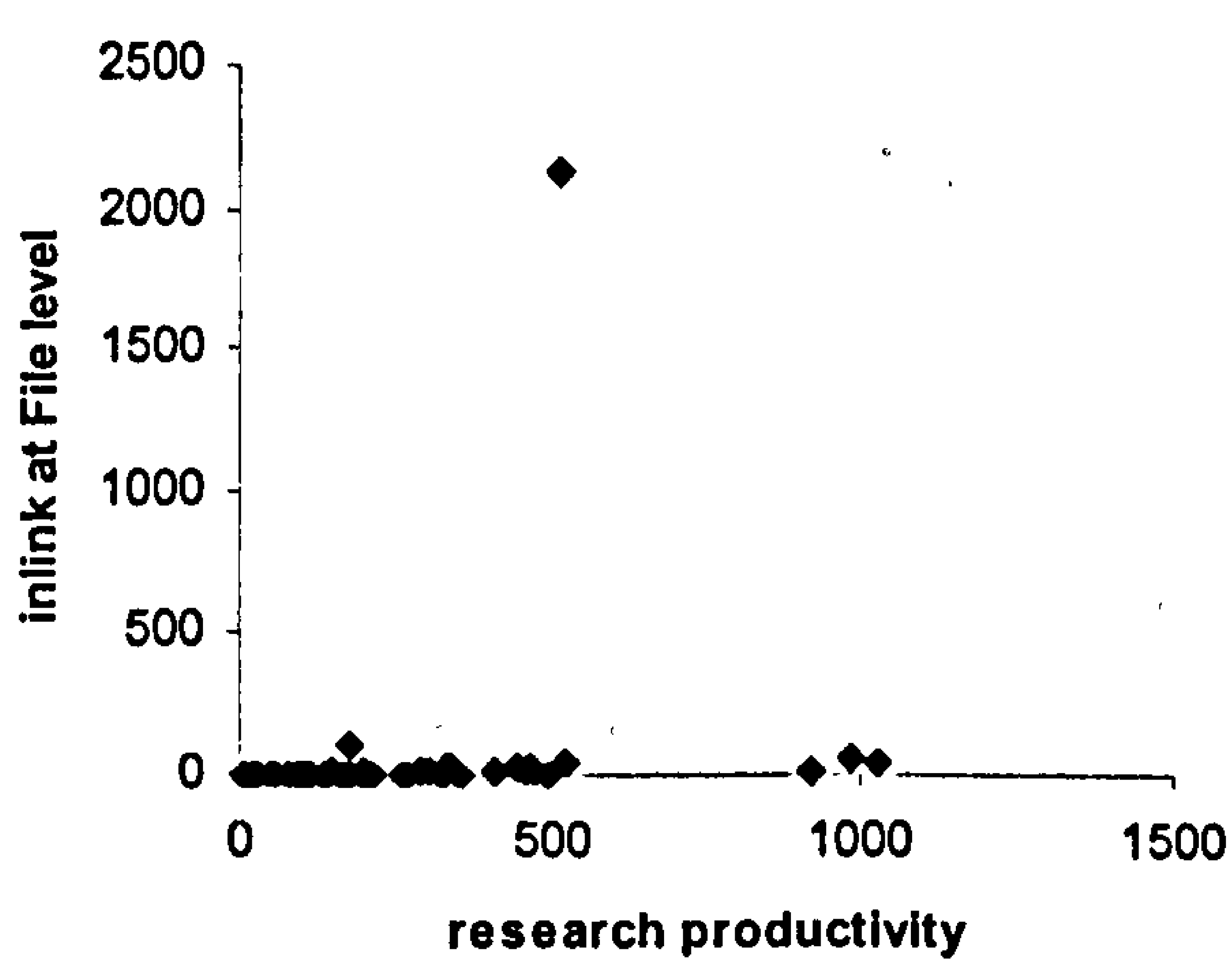


Figure 6.33 File inlinks against research productivities for UK biology departments

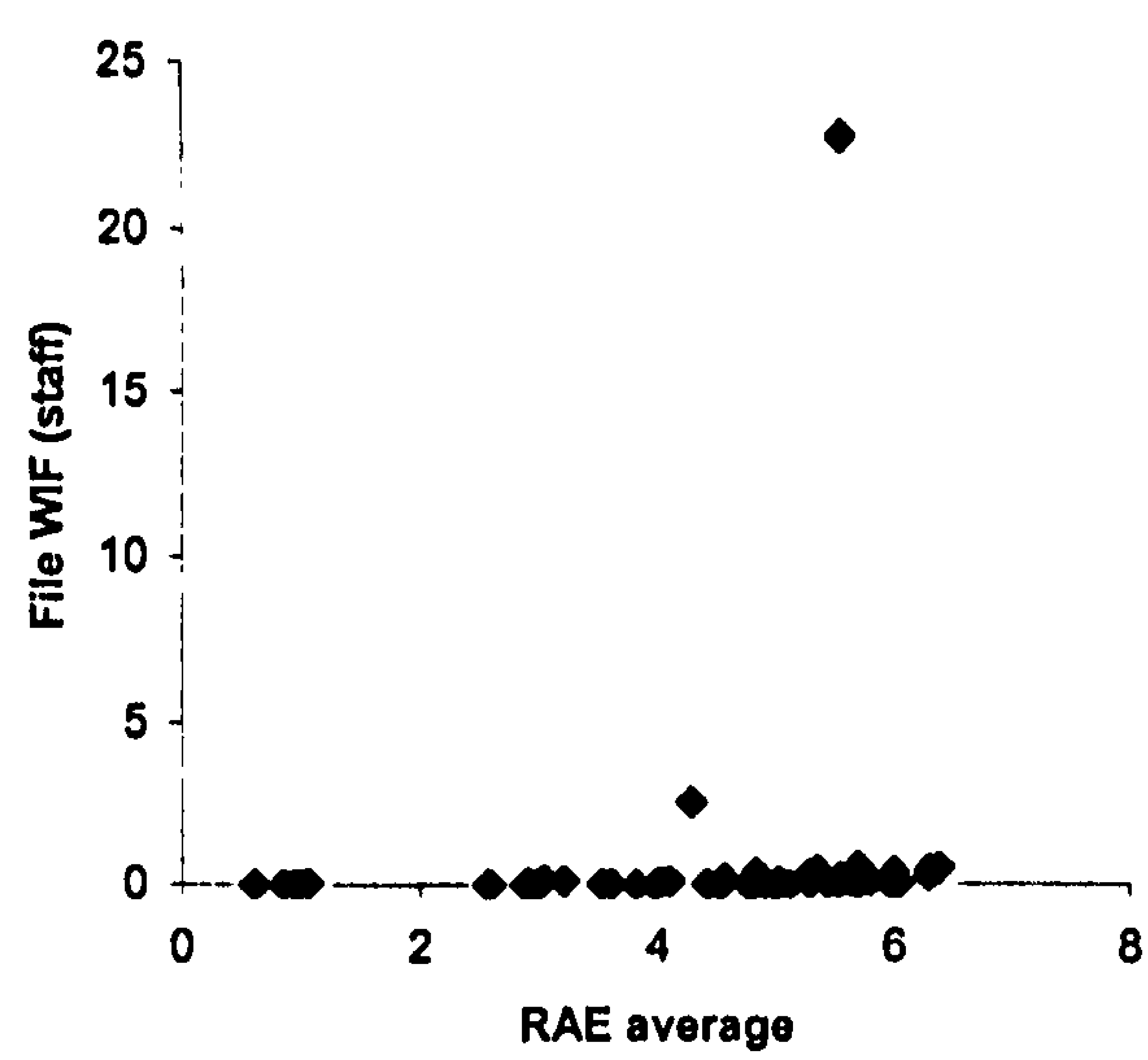


Figure 6.35 File WIFs (staff) against RAE averages for UK biology departments

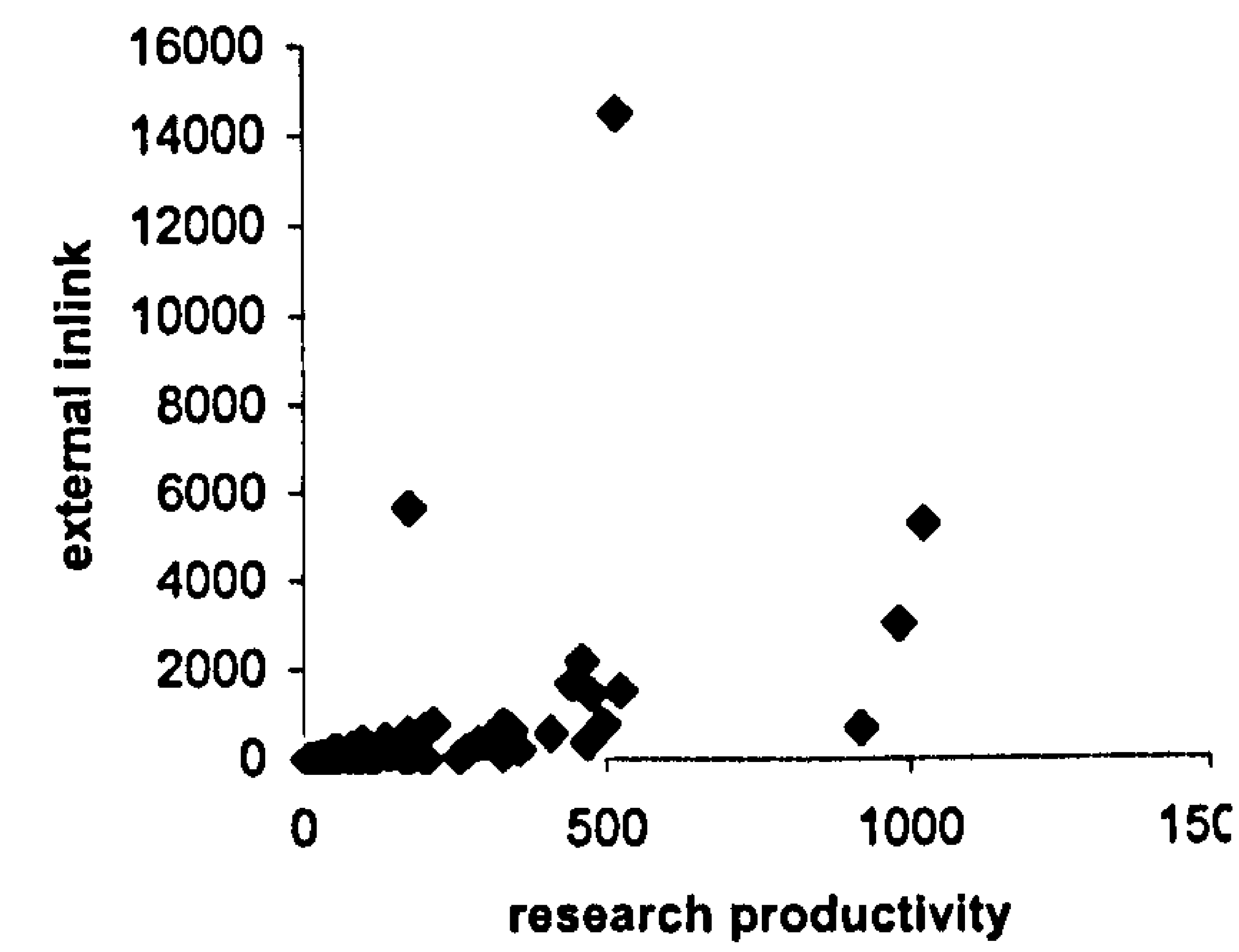


Figure 6.34 External inlinks against research productivities for UK biology departments

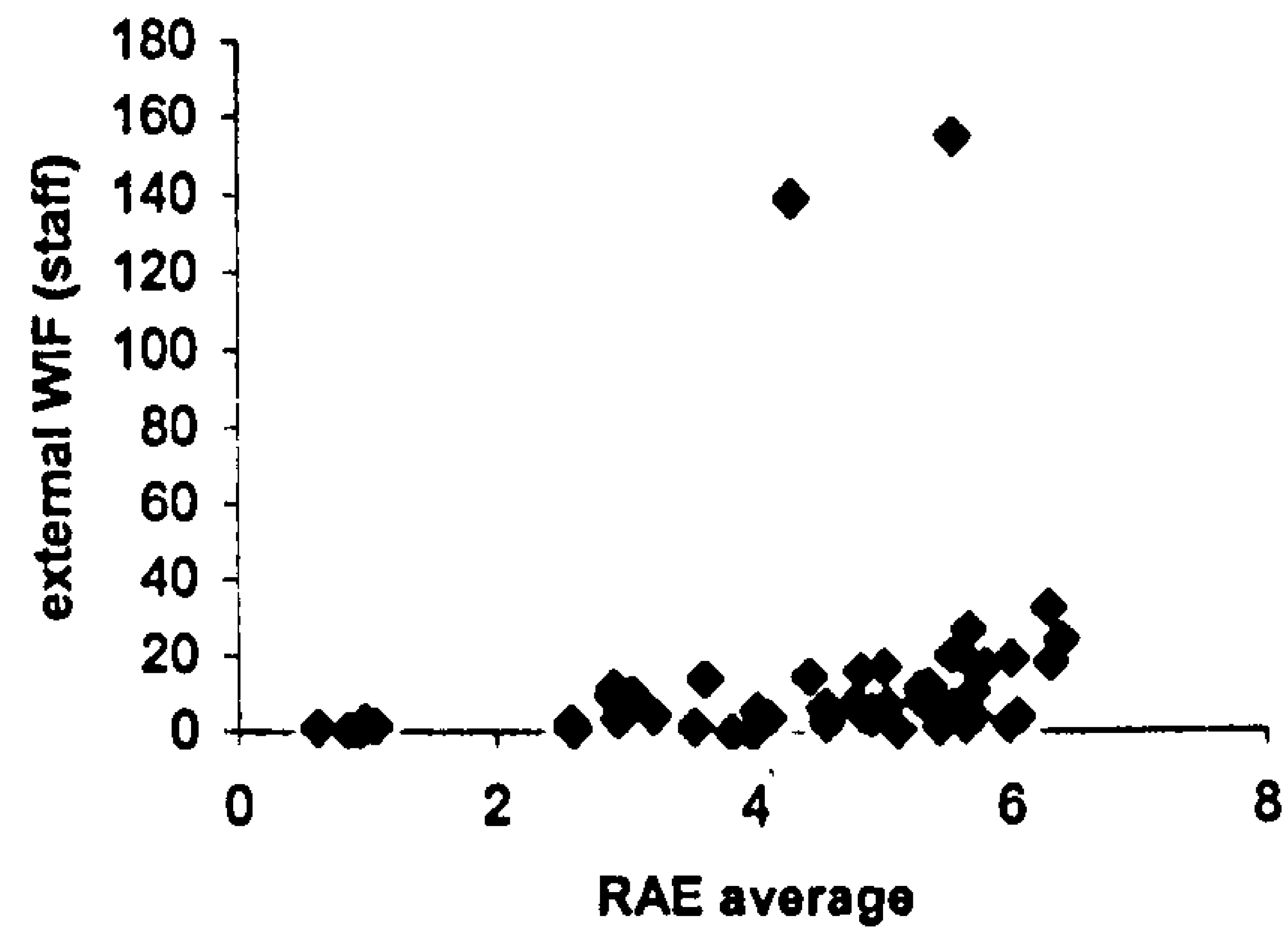


Figure 6.36 External WIFs (staff) against RAE averages for UK biology departm

6.1.4 Correlation Tests for Canadian and UK Departments' Citation Counts

6.1.4.1 Results for the Canadian Departments between Citations and Research Grants

Tables 6.61 and 6.62 illustrate the Pearson and Spearman correlations between the citation counts and research grants for Canadian departments. Apart from the values between CPP and research grants per staff member for biology and chemistry are not significant, the rest are significant at the 1% value.

Table 6.61 Pearson correlations (* = significant at the 5 % level, ** = significant at the 1% level)

Data sets	Biology (n=40)	Chemistry (n=39)	Physics (n=37)
Citations versus research grants	0.709**	0.876**	0.704**
Citations per staff member versus research grants per staff member	0.379*	0.716**	0.633**
CPP versus research grants per staff member	0.266	0.085	0.546**

Table 6.62 Spearman correlations (* = significant at the 5 % level, ** = significant at the 1% level)

Data sets	Biology (n=40)	Chemistry (n=39)	Physics (n=37)
Citations versus research grants	0.853**	0.888**	0.864**
Citations per staff member versus research grants per staff member	0.484**	0.740**	0.622**
CPP versus research grants per staff member	0.355*	0.238	0.662**

6.1.4.2 Results for the UK Departments between Citations and RAE Ratings

Tables 6.63 and 6.64 illustrate the correlations between citation counts and RAE ratings for the UK departments. Nearly all values are significant at the 1% level.

Table 6.63 Pearson correlations (* = significant at the 5 % level, ** = significant at the 1% level)

Data sets	Biology (n=53)	Chemistry (n=41)	Physics (n=44)
Citations versus RAE research productivities	0.857**	0.847**	0.640**
Citation per staff member versus RAE averages	0.550**	0.626**	0.351*
CPP versus RAE averages	0.620**	0.554**	0.376*

Table 6.64 Spearman correlations (* = significant at the 5 % level, ** = significant at the 1% level)

Data sets	Biology (n=53)	Chemistry (n=41)	Physics (n=44)
Citations versus RAE research productivities	0.903**	0.758**	0.739**
Citation per staff member versus RAE averages	0.731**	0.561**	0.332*
CPP versus RAE averages	0.650**	0.526**	0.345*

6.2 Results of the Target Page Classification

6.2.1 Disappearance Rate for Each Set of Departments

Table 6.65 lists the proportion of 'disappeared' pages for each set of departments in each country. The 'disappeared' proportion for Australian chemistry target pages is high at 0.355. The school of chemistry in University of Sydney attracted 29 links from its peers, 12 had already disappeared when those target pages were visited. The major reason is the reorganization of homepages for staff members. Without these pages, the 'disappeared' proportion for Australian chemistry departments could be 0.226. All target pages were visited in May 2004. Even though the link data was collected in June 2003 for the UK, June-July 2003 for Australia and January 2004 for Canada, the 'disappeared' rates are very similar. This shows that the academic web pages in the three countries are relatively stable. On the other hand, some of the 'disappeared' target pages may already have gone during the crawling of the link data, as a lot of old domains or URLs for the departments are included. If this is the case, the links that pointed to those 'disappeared' pages were just not updated in time. In order to get a more complete result, it would be better to visit the target pages immediately after the crawling.

Table 6.65 Proportions of disappeared target pages

Country	Biology	Chemistry	Physics
Australia	0.211	0.355	0.254
Canada	0.208	0.114	0.198
UK	0.081	0.175	0.193

6.2.2 Target Page Classification for Australian Departments

Table 6.66 displays the number of occurrences for target pages in different groups and relevant proportions for biology, chemistry and physics departments in Australia. Figure 6.37 illustrates the different proportions of those target pages in different categories for those three types of departments. Some special target pages for Australian physics departments are listed in appendix H1. For the target pages about conference information, one is about Visa applications for foreign participants, the other is to help research students to get funding to attend a conference. They are academic related, but without containing any subject content. However, they are relevant information for the source pages.

Table 6.66 Different categories of target pages for Australian biology, chemistry and physics departments

Category	Biology		Chemistry		Physics	
	number	proportion	number	proportion	number	proportion
Department	20	0.222	20	0.215	146	0.174
Group	15	0.167	6	0.065	136	0.162
Staff	21	0.233	20	0.215	79	0.094
Resource	15	0.167	12	0.129	263	0.314
Disappeared	19	0.211	33	0.355	213	0.254
Others	0	0	2	0.021	0	0

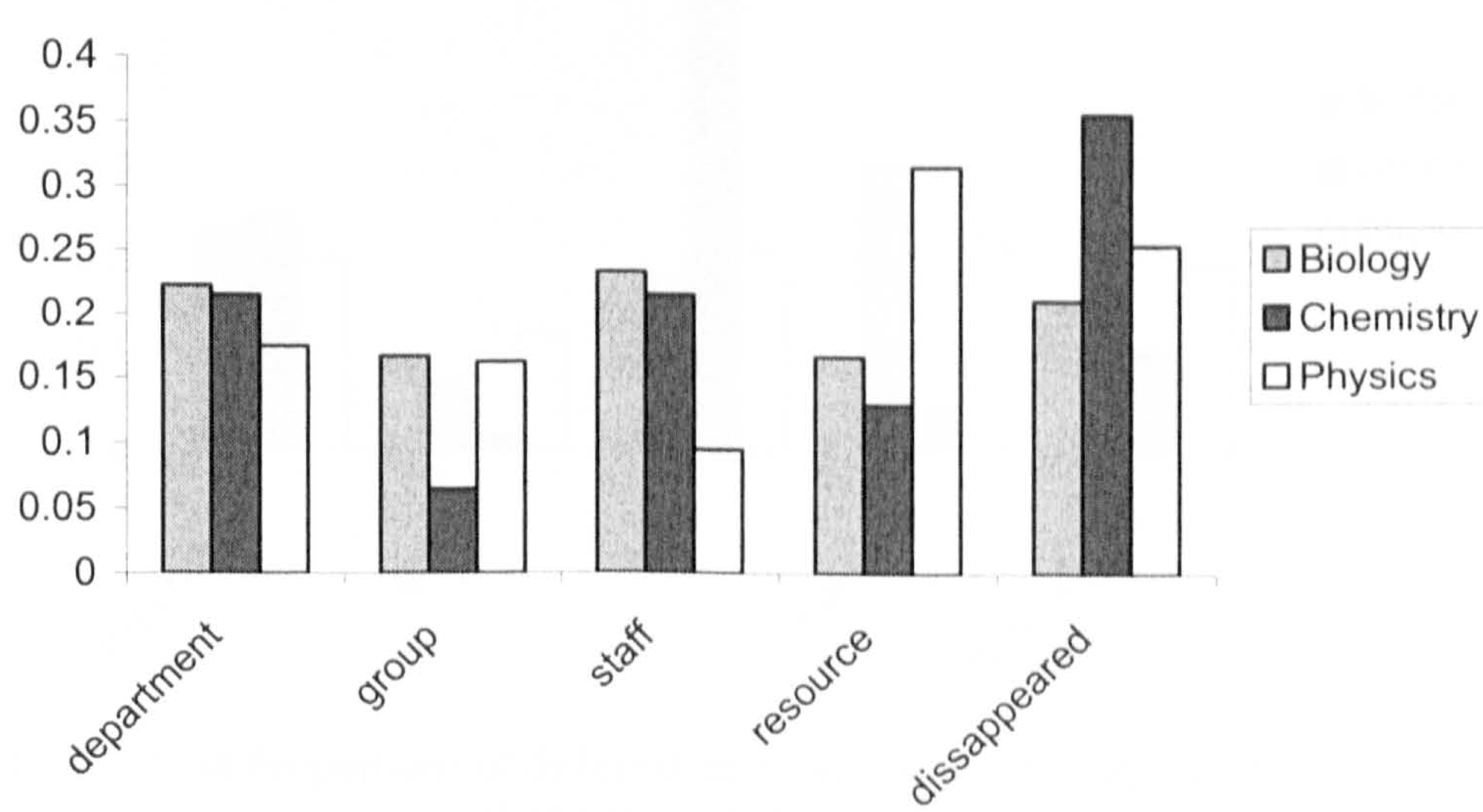


Figure 6.37 Proportions of different categories of target pages for Australian biology, chemistry and physics departments

6.2.3 Target Page Classification for Canadian Departments

Table 6.67 displays the number of web pages grouped in each category and relevant proportions for Canadian biology, chemistry and physics departments. Figure 6.38 illustrates the proportions of different categories of target pages for Canadian biology, chemistry and physics departments. The special target pages in each set of departments in Canada are listed in Appendixes H2 to H4.

A web page from the chemistry department of the University of Calgary (http://www.cobalt.chem.ucalgary.ca/ziegler/Lec.chm373/mathtools.html) hosts mathematics tools (Mathematical Basics and Mathematics for Quantum Mechanics), which are created by Dan Thomas in the chemistry department of the University of Guelph. Each page contains a link back to its author's homepage (http://www.chembio.uoguelph.ca/thomas/thomas.htm), altogether there are 392 of this type of link. The total number of links among Canadian chemistry departments is 1,057. This is the reason why the proportion of staff target is high at 0.571.

Table 6.67 Different categories of target pages for Canadian biology, chemistry and physics departments

Category	Biology		Chemistry		Physics	
	number	proportion	number	proportion	Number	proportion
Department	86	0.232	262	0.248	237	0.201
Group	27	0.073	18	0.017	150	0.128
Staff	69	0.186	604	0.571	250	0.213
Resource	112	0.302	50	0.047	272	0.231
Disappeared	77	0.208	121	0.114	233	0.198
Others	0	0	2	0.002	34	0.029

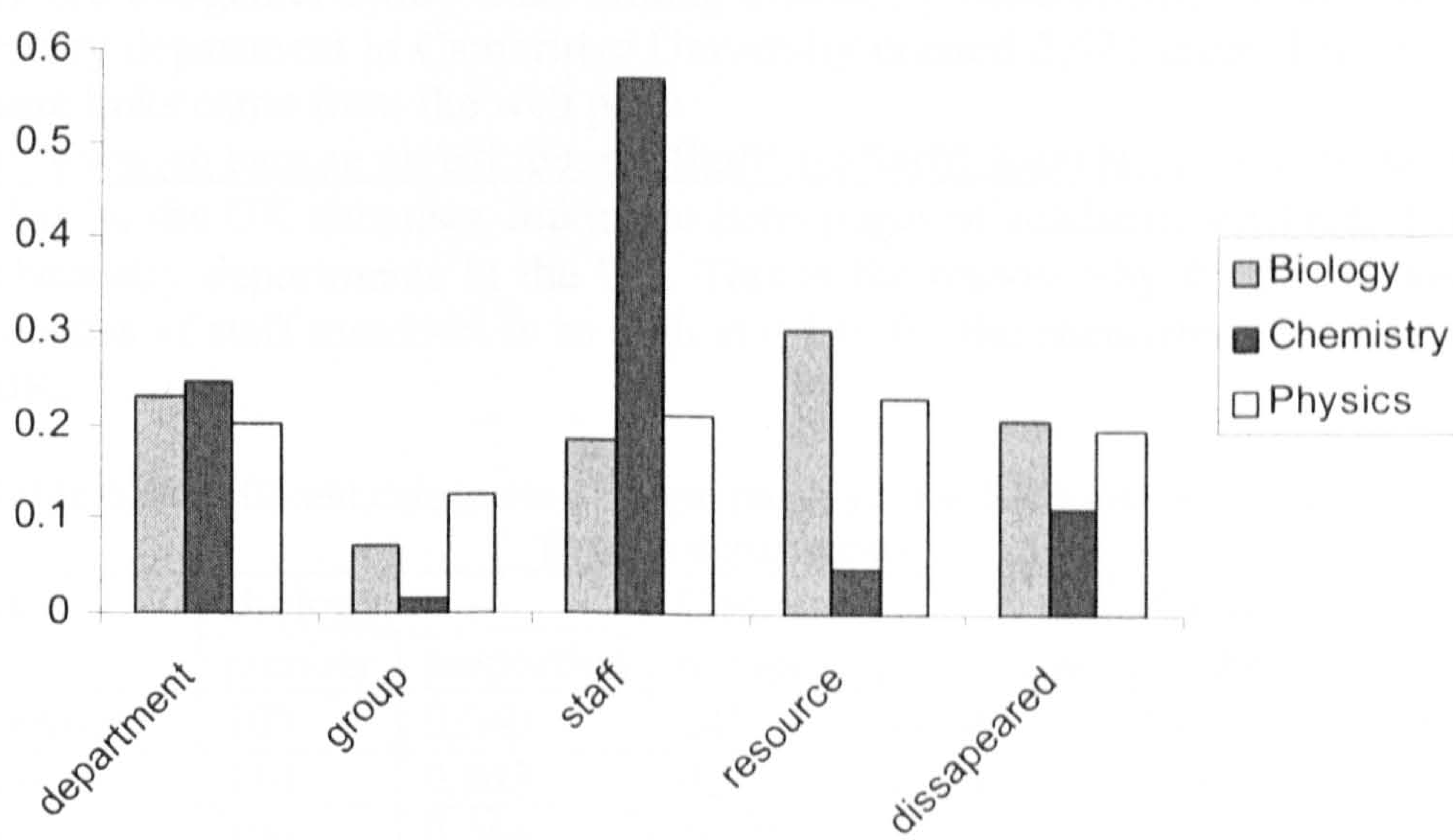


Figure 6.38 Proportions of different categories of target pages for Canadian biology, chemistry and physics departments

6.2.4 Target Page Classification for UK Departments

Table 6.68 displays different categories of target pages for the UK biology, chemistry and physics departments. Figure 6.39 illustrates proportions of different categories of target pages for the UK biology, chemistry and physics departments. Some special target pages identified for departments in the UK are listed in Appendixes H5 to H7.

The student homepage (http://student.cryst.bbk.ac.uk/~esodh01/) found for the biology department in Birkbeck University was the only student's homepage found in this study. The source page pointing to the student's homepage was visited (http://www.biochem.ucl.ac.uk/~sodhi/). By the time it was visited, the page had already disappeared. The target page has also a link to the source page. The student might study in the two universities. He/She may create one web page in each of the two universities and create links pointing to each other.

There are altogether 2,676 links among the biology departments in the UK. The biology department in the Cambridge University links to the biochemistry and molecular biology department in the University College London 1,748 times. 1,720 link to the databases hosted in that department. Details of those databases are listed in Appendix H7. This is the reason why the proportion of resources for UK biology departments is so high at 0.767.

In Appendix H6, apart from some online journals, papers, student societies and national or international organizations, some non-academic related target pages were identified, such as maps, server statistics, history and phone numbers of the departments. They are categorised as resources because they are general resources. Even though they are not academically related, they are related to the department, and are more serious than recreational web pages.

There are altogether 3,864 links among chemistry departments in the UK. The chemistry department in Cambridge University created 2,570 links. The majority of these links came from the web page http://www.ch.cam.ac.uk/c2k/people/StaffList/StaffList00.html which hosts a link list for the UK chemists, linking to homepages of academic staff members in the chemistry departments in the UK. This is the reason why the proportion for homepages of staff members is so high at 0.501 for the chemistry departments in the UK.

Table 6.68 Different categories of target pages for the UK biology, chemistry and physics departments

Category	Biology		Chemistry		Physics	
	number	proportion	number	proportion	number	proportion
Department	107	0.040	247	0.064	202	0.189
Group	114	0.043	92	0.024	256	0.239
Staff	176	0.066	1937	0.501	91	0.085
Resource	2054	0.767	893	0.231	303	0.284
Disappeared	217	0.081	675	0.175	206	0.193
Others	8	0.003	20	0.005	11	0.010

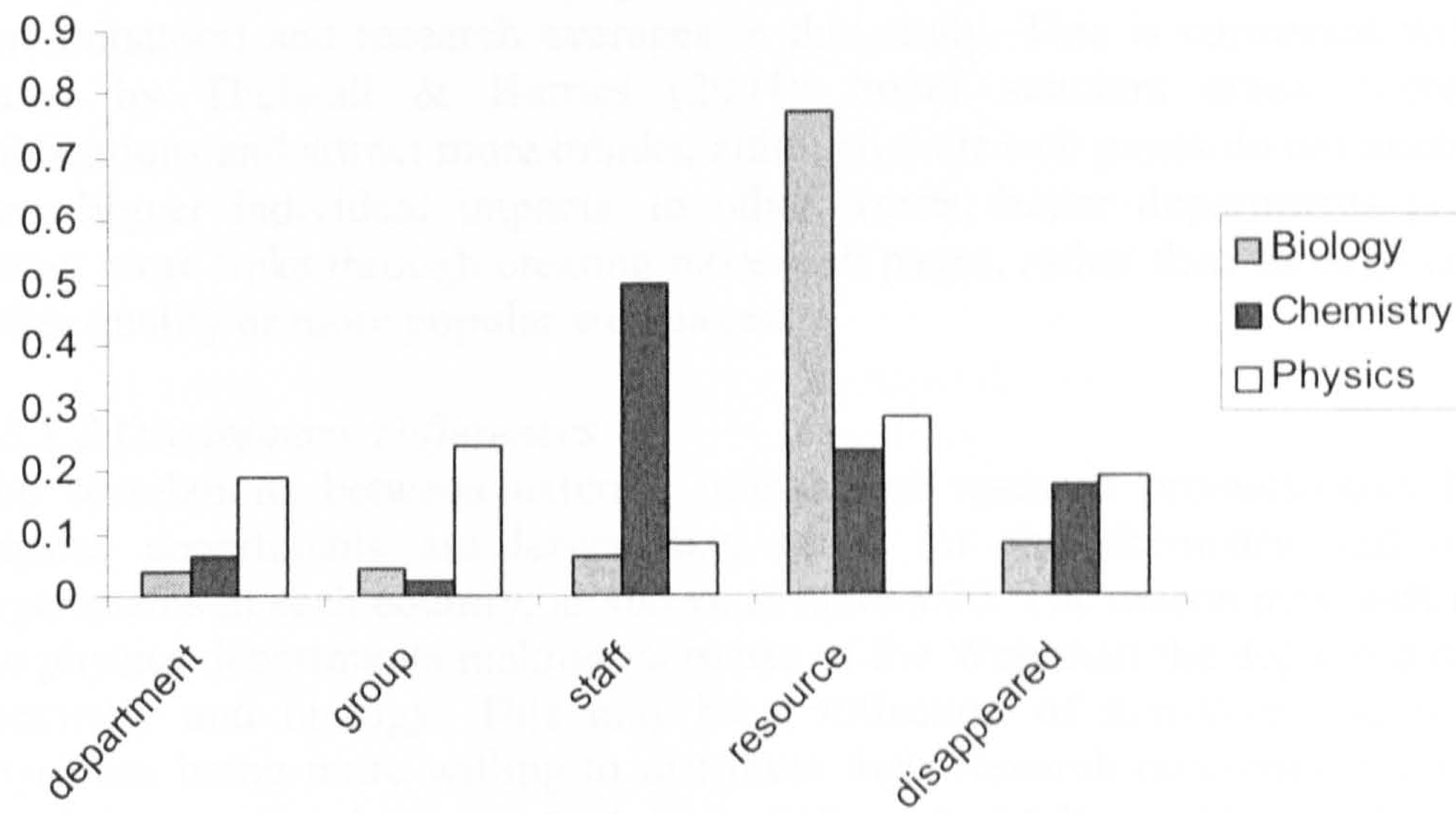


Figure 6.39 Proportions of different categories of target pages for the UK biology, chemistry and physics departments

6.3 Discussion

6.3.1 The Significance of the Correlation Tests

The proportions of departmental inlinks are small with respect to the number of university inlinks among the same set of departments, as shown in table 6.69. Despite this, the correlations are nearly all significant, at the 1% level, between the number of inlinks and research productivity for the physics, chemistry and biology departments. This gives evidence that the three types of departments publish enough on the Web to conduct meaningful link investigations, at least in these three countries.

Table 6.69 Proportions of departmental inlinks with respect to external links to universities which host the set of departments from the same set. UL stands for inlinks to universities that host the departments; DL stands for inter-departmental inlinks; P stands for the proportion of inter-departmental inlinks.

Department	Australia			Canada			UK		
	UL	DL	P	UL	DL	P	UL	DL	P
Biology	381	90	0.236	2053	371	0.181	11515	2676	0.232
Chemistry	402	93	0.231	1999	1057	0.529	7072	3864	0.546
Physics	2085	837	0.401	2333	1176	0.504	10021	1069	0.107

6.3.1.1 Size Effects

It may be argued that the significant correlations are caused by the size effects, with bigger departments having more faculty members; creating more web pages and hyperlinks; conducting more research; and attracting more inlinks. After the size effects of the departments are removed by dividing the number of links by the number of academic staff members (i.e. moving to WIFs and research averages), the correlations are lower but many are still statistically significant. This shows that the significant correlations not only caused by departmental size, although it is one of the reasons. There are hardly any significant correlations between WIFs (with web pages, directories, domains or departments as denominators) and research averages in this study. This is consistent with that found by Thelwall & Harries (2004): better scholars create more web publications and attract more inlinks, although their web pages do not necessarily have bigger individual impacts. In other words, better departments probably attract more links through creating more web pages, rather than through creating better quality or more popular web pages.

6.3.1.2 Disciplinary Differences

The correlations between external inlinks and research productivities for the physics departments are larger than those for the chemistry and biology departments in each country, as shown in table 6.70. The reason may well be that the physics departments making more use of the Web than the departments from chemistry and biology. This may be a reflection of non-Web phenomenon; physicists being more willing to distribute their research outcomes quickly and widely than chemists and biologists (Kling & McKim, 2000). The largest correlation coefficient value is for the Australian physics departments.

Table 6.70 Pearson correlations between external inlinks (from AltaVista) and research productivities (n is the number of departments).

Department	UK	Australia	Canada
Physics	0.647** (n = 44)	0.900** (n = 24)	0.766** (n = 37)
Chemistry	0.642** (n = 41)	0.410* (n = 25)	0.601**(n = 40)
Biology	0.442** (n = 53)	0.573** (n = 26)	0.497** (n = 40)

6.3.1.3 The Effects of Alternative Document Models

The application of ADMs in this project does not give the significantly differing results for departments that it does for whole universities. Nevertheless, it is very useful for removing the effects of apparent anomalies. An example is the UK biology departments; the domain and department document models effectively remove the anomaly caused by the biology department of the University College London, which hosts a lot of databases, and attracts large number of inlinks from its peers in the UK, as shown in table 6.57. Another example is the Canadian chemistry departments; the directory, domain and department document models successfully remove the anomaly caused by the chemistry department of the University of Calgary, which mirrors a set of lecture notes created by a staff member from the University of Guelph, and each page of the lecture notes has a link back to its creator's homepage, as shown in table 6.25. The domain document model does not give the most significant correlation coefficient values for the departments, in contrast to the case for universities. The reason may well be that some departments' urls are organized by different domain names, while others are by different directories. Universities with different departments which exploit different url conventions may average this effect out. Nevertheless, the ADMs have been confirmed to be an efficient technique to remove apparent anomalies from departmental link counts. It is therefore desirable to use the ADMs in future departmental link investigations.

6.3.1.4 The Correlation Tests for the Citation Counts

Section 6.1.4 reports the correlation coefficient values between the citation counts and relevant research measures for the UK and Canadian departments. The fact that nearly all values are significant at the 1% level validates the semi-automatic citation counting technique proposed in this study.

6.3.2 The Target Page Types

The sparseness of journal papers as target pages means that the links between departments are not the same as citations between journal papers. However, the overwhelmingly academically related content in the target pages may suggest that departmental interlinking reflects informal scholarly communication amongst peers. Furthermore, the fact that no recreational target pages were found in this project suggests that departmental interlinking is more consistently serious than that for whole universities.

In the three countries, the physics departments target the 'resources' type of web pages more than any other type. Job information can be found in the target pages for all the three country's physics departments. This suggests that physics departments make more use of the Web than biology and chemistry departments.

As each department follows different conventions for creating web pages and hyperlinks, it is hard to find the exact motivation for departmental interlinking. A set of departments' linking behaviour is especially vulnerable to an individual department's linking behaviour. In order to avoid certain departments rendering the whole research meaningless, isolated anomalies should be removed by applying a technique such as the ADMs.

7. National and Disciplinary Linking Differences

In this chapter, the results with respect to the four different aspects, which are designed in section 4.2.3, are reported in tables 7.1 to 7.15. Figures 7.1 to 7.28 are then drawn from the results, and reorganised both along country and disciplinary lines to facilitate pattern seeking. When a comparison is made along country lines, it is within the same discipline. Comparisons are also made within each country along disciplinary lines.

7.1 Results from the Four Different Aspects

7.1.1 General Web Uses

7.1.1.1 Median Number of Web Pages

Table 7.1 lists the median number of web pages (from SocSciBot) for each of the nine sets of departments. In comparison, the sizes of departments' web sites in the UK are the largest amongst the three countries, while those in Australia are the smallest.

Table 7.1 Median web pages

Country	Physics	Chemistry	Biology
Australia	327.5	220	267
Canada	568	279.5	584.5
UK	820	426	576

7.1.1.2 Median Number of External Outlinks

Table 7.2 lists the median number of external outlinks (from SocSciBot) for each of the nine sets of departments. Generally, the UK departments make the largest number of outlinks, while Australian departments make the smallest. The only exception is that Canadian chemistry departments' external outlinks is marginal smaller than those from Australian chemistry departments.

Table 7.2 Median external outlinks

Country	Physics	Chemistry	Biology
Australia	213.5	119	71.5
Canada	562	118	419.5
UK	595.5	182	375

7.1.1.3 Median External Inlinks from the Whole Web

Table 7.3 lists the median number of inlinks from the whole Web (from AltaVista) for each of set of departments. UK departments attract the largest number of inlinks from the whole Web, while Australian departments attract the smallest, with two exceptions from Canadian physics and chemistry departments.

Table 7.3 Median external inlinks from the whole Web

Country	Physics	Chemistry	Biology
Australia	101.5	100	80.5
Canada	467	87.5	102.5
UK	238	176	219

7.1.1.4 Proportion of National Inlinks

The proportion of inlinks for each set of departments from its own country domain with respect to those from the whole Web is listed in table 7.4. The higher proportion of national inlinks for a set of departments, the less visible is the set of departments from the rest of the Web. The proportions of Australian chemistry and biology departments are larger than other sets of departments. They are more visible in Australia than from the whole Web. The proportion of chemistry departments in each country is larger than those of physics and biology departments. The chemistry departments' web sites are more nationally visible at least in these three countries.

Table 7.4 Proportion of national inlinks

Country	Physics	Chemistry	Biology
Australia	0.17	0.61	0.40
Canada	0.21	0.31	0.20
UK	0.22	0.26	0.26

7.1.2 National Peer Inlinks

7.1.2.1 Interconnection Rates

The interconnection rates listed in table 7.5 below shows how much each set of departments is interconnected. The results are calculated from the link data collected by SocSciBot3. Overall, the interconnection rates for biology departments are relatively low, while those of physics departments are high. UK physics departments are an exception. The interconnection rate for the set of departments is smaller than that of the chemistry departments, but is still larger than that of the biology departments though.

Table 7.5 Interconnection Rates

Country	Physics	Chemistry	Biology
Austalia	0.21	0.067	0.074
Canada	0.25	0.135	0.108
UK	0.131	0.19	0.067

7.1.2.2 Median National Peer Inlinks

Table 7.6 lists the median number of national peer inlinks for each set of departments. Chemistry departments in the UK attract the most inlinks from their national peers. Apart from the UK physics departments, each set of physics departments attract the largest number of national peer inlinks, while those for biology departments are the smallest.

Table 7.6 Median number of inlinks from national peer departments

Country	Physics	Chemistry	Biology
Australia	14	2	1
Canada	16	8.5	3
UK	4.5	82	3

7.1.3 International Peer Interlinking

Although the international peer inlinks are small in number, as shown in appendix G1, interesting information can be identified along both disciplinary and country lines.

7.1.3.1 Adapted Mean International Peer Inlinks

Table 7.7 lists the adapted mean international peer inlinks for each set of departments. The values for the UK chemistry and biology are the largest in the three countries, followed by Australia and then Canada. The physics departments in Australia and Canada have the largest mean international peer inlinks along disciplinary lines, except those in the UK.

Table 7.7 Adapted mean international peer inlinks for each set of departments

Coutry	Physics	Chemistry	Biology
Australia	0.175	0.053	0.025
Canada	0.100	0.038	0.021
UK	0.079	0.081	0.050

7.1.3.2 Link Propensities

The link propensities are calculated both with the product of staff members and web pages as denominators. Since the link propensities are extremely small, as listed in appendix G2, the values in tables 7.8 and 7.9 are normalized by setting the largest value in each table as 100. From table 7.8, in each country the physics departments tend to attract the most links, followed by chemistry and then biology. The exceptions are: from Canada to the UK, where the chemistry departments tend to attract more links than the physics departments; and from Australia to Canada, the chemistry departments tend to attract fewer links than the biology departments. In table 7.9, the link propensities for the chemistry departments are the largest for each country. This may be caused by the smallest web sites for the chemistry departments in each country, as shown in table 7.1. So the results in table 7.8 are more reliable than those in table 7.9.

For the physics departments, the UK tends to link to Canada and Australia, while Australia tends to link to Canada. For chemistry departments, Canada and the UK tend to link to Australia, and Canada tends to link to the UK. For the biology departments, Canada and Australia tend to link to the UK, and Australia tends to link to Canada, as shown in table 7.8.

Table 7.8 Link propensities between sets of departments in same discipline from Australia, Canada and the UK (with product of staff members as denominators)

Country from to	Physics	Chemistry	Biology
Canada to UK	19.8	56.7	9.76
UK to Canada	44.6	37	2.3
Australia to UK	70.4	39.5	7.84
UK to Australia	100	41	3.78
Australia to Canada	81.1	2.69	4.28
Canada to Australia	39.7	14.5	4

Table 7.9 Link propensities between sets of departments in same discipline from Australia, Canada and the UK (with product of web pages as denominators)

Country from to	Physics	Chemistry	Biology
Canada to UK	5.8	77.4	14.9
UK to Canada	13.1	50.4	3.5
Australia to UK	19.1	96.7	14.4
UK to Australia	27.1	100	7.1
Australia to Canada	21.4	7.9	33.1
Canada to Australia	10.5	43	31

7.1.4 Interactions with Different Top Level Domains

7.1.4.1 Proportions of External Inlinks from Different Top Domains

Tables 7.10 to 7.12 list proportions of inlinks from different domains for each of the nine sets of departments. For each set of departments, the proportion of inlinks from com domain is larger than that of edu domain, while proportions of edu.au, ac.uk inlinks are larger than those of com.au, co.uk respectively, as shown in tables 7.10 and 7.12.

Table 7.10 Proportions of inlinks from different domains for Australian departments

Domain	Physics	Chemistry	Biology
com	0.229	0.093	0.137
edu	0.127	0.052	0.230
au	0.174	0.611	0.401
org	0.077	0.020	0.066
net	0.072	0.013	0.023
com.au	0.023	0.017	0.034
edu.au	0.112	0.556	0.288

Table 7.11 Proportions of inlinks from different domains for Canadian departments

Domain	Physics	Chemistry	Biology
com	0.197	0.200	0.253
edu	0.161	0.149	0.184
ca	0.205	0.313	0.195
org	0.080	0.048	0.099
net	0.043	0.030	0.043

Table 7.12 Proportions of inlinks from different domains for the UK departments

Domain	Physics	Chemistry	Biology
com	0.141	0.216	0.147
edu	0.095	0.131	0.137
uk	0.221	0.260	0.259
org	0.059	0.055	0.064
net	0.039	0.033	0.028
co.uk	0.033	0.028	0.037
ac.uk	0.171	0.221	0.200

7.1.4.2 Proportions of External Outlinks to Different Domains

Tables 7.13 to 7.15 list the proportions of links to different domains from each set of departments with respect to the total number of external outlinks made. The results are calculated by the link data collected by the SocSciBot3. Appendix F lists the numbers of oulinks in detail. With regard to the three country domains, uk, au and ca, each set of departments tends to link to its own country domain first. Then the departments in Australia and Canada tend to link more to the uk domain. Departments in the UK tend to link to the au more than the ca domain. This is in parallel with the collaboration patterns found through co authorship analysis in bibliometrics (Glänzel, 2001; Glänzel & Schubert, 2001). The UK is the collaboration centre in formal scholarly communication in the world, and it collaborates more with Australia than with Canada.

All the three types of departments in Australia are found to outlink heavily to the com domain, compared with those from the UK and Canada. This may suggest that the departments in Australia cooperate more with industry.

Table 7.13 Proportions of links to domains for the Australian departments

Domain	Physics	Chemistry	Biology
uk	0.042	0.050	0.035
au	0.166	0.224	0.314
ca	0.011	0.008	0.011
edu	0.137	0.090	0.111
com	0.357	0.400	0.312
gov	0.042	0.023	0.030
org	0.109	0.091	0.064
net	0.017	0.016	0.023

Table 7.14 Proportions of links to different domains for Canadian departments

Domain	Physics	Chemistry	Biology
uk	0.033	0.048	0.047
au	0.011	0.008	0.005
ca	0.132	0.206	0.194
edu	0.221	0.169	0.159
com	0.253	0.286	0.250
gov	0.109	0.032	0.098
org	0.132	0.127	0.143
net	0.016	0.019	0.014

Table 7.15 Proportions of links to different domains for theUK departments

Domain	Physics	Chemistry	Biology
uk	0.161	0.138	0.439
au	0.013	0.014	0.005
ca	0.013	0.011	0.004
edu	0.204	0.158	0.085
com	0.150	0.120	0.115
gov	0.050	0.167	0.075
org	0.181	0.125	0.099
net	0.018	0.006	0.008

7.2 Link Differences Along National and Disciplinary Lines

7.2.1 National Differences

7.2.1.1 General Web Uses

1. Median web pages. As shown in Figure 7.1, the Australian physics, chemistry and biology departments' median numbers of web pages are all the smallest amongst the three countries. Except that the UK biology departments' median number of web pages is marginally smaller than that for the Canadian biology departments, the UK physics and chemistry departments' median numbers of web pages are the largest.

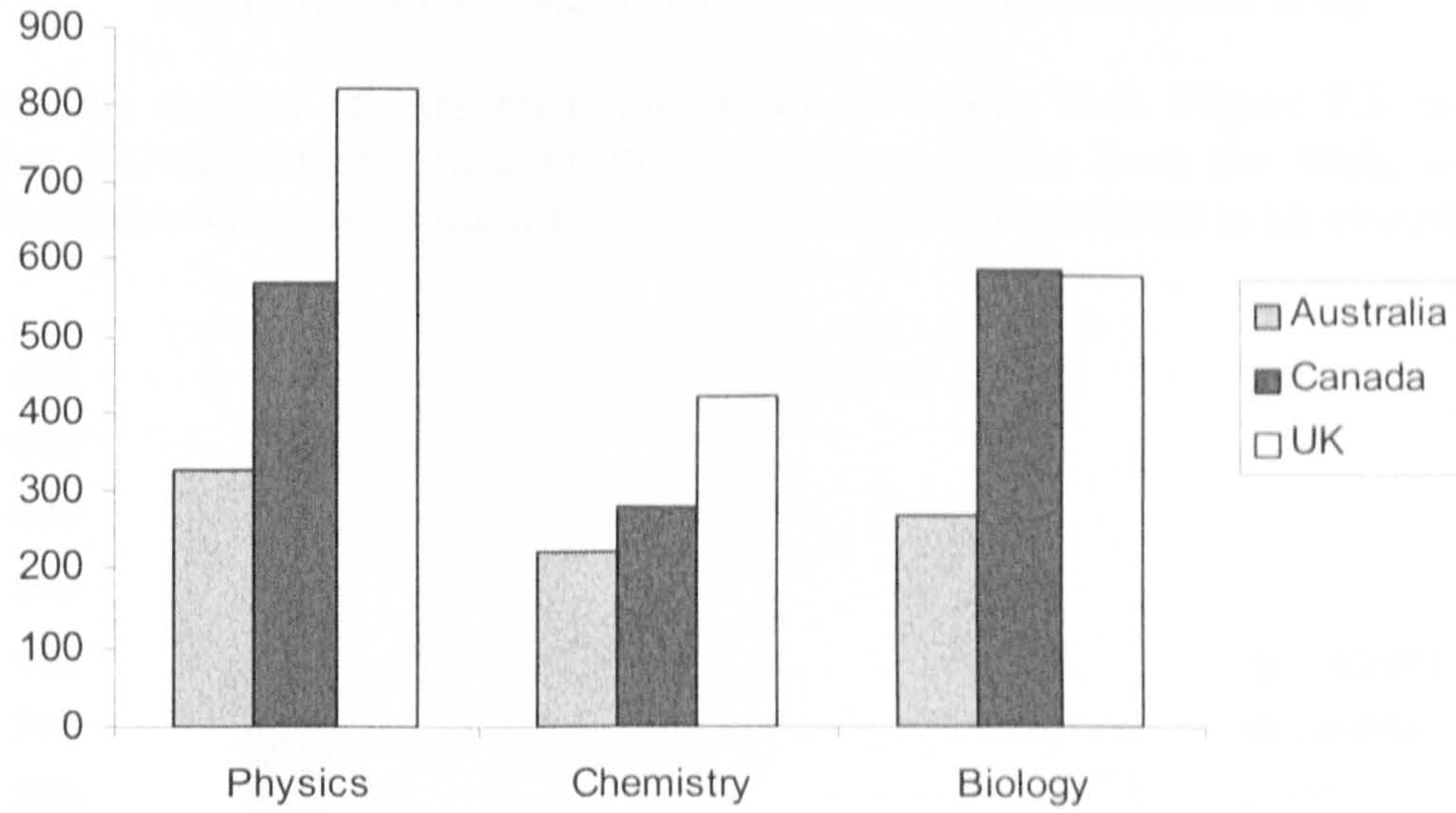


Figure 7.1 Median web pages by country (SocSciBot3 data)

2. Median external outlinks. The median external outlinks for Australian departments are the smallest in all three disciplines, as shown in figure 7.2. For the UK, apart from the biology departments, the median external outlinks for the physics and chemistry departments are both the largest amongst the three countries.

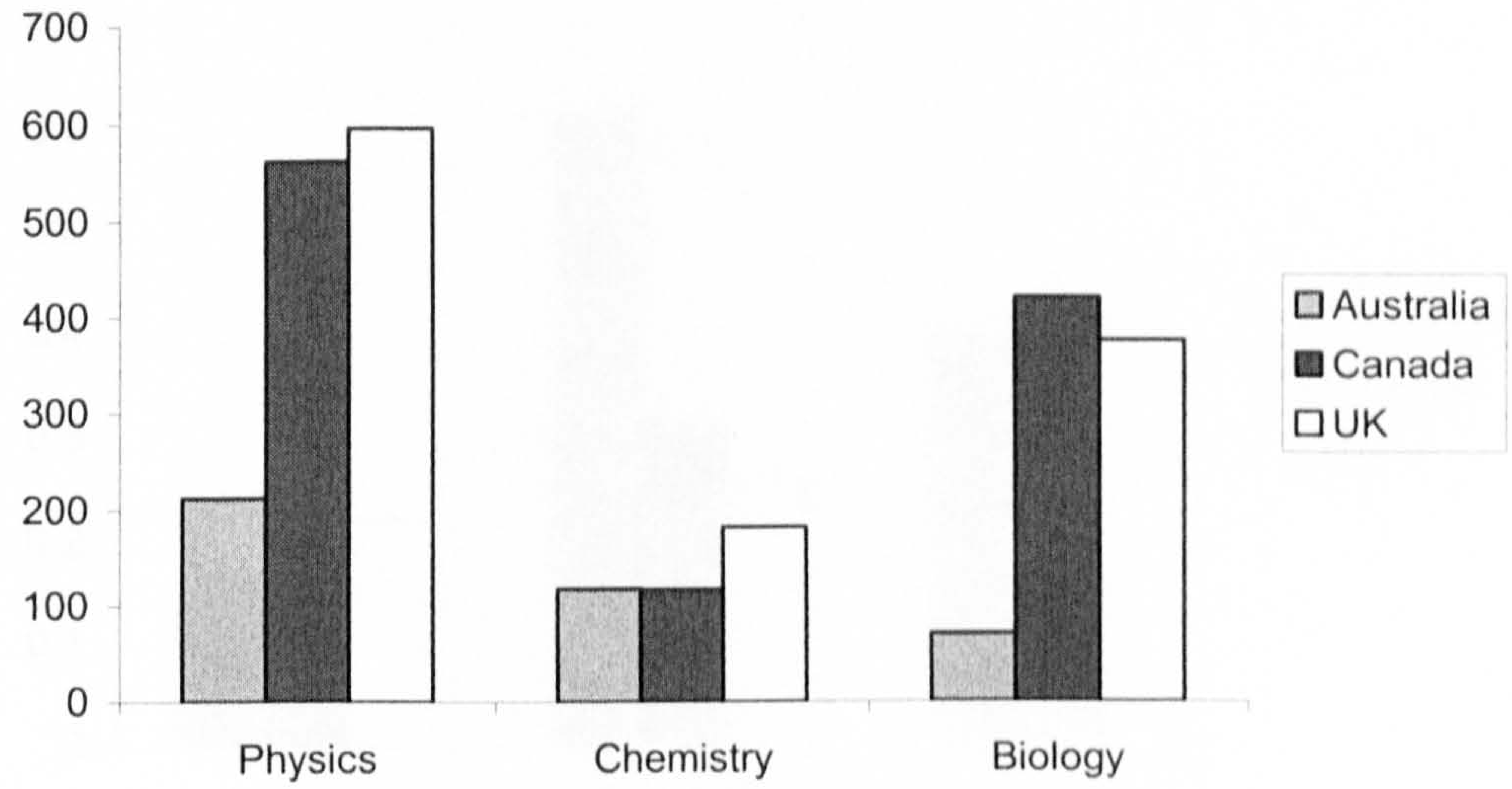


Figure 7.2 Median external outlinks by country (SocSciBot3 data)

3. Median number of external inlinks from the whole Web. Figure 7.3 shows that UK departments tend to attract the most inlinks from the Web, while Australia departments the least. Canadian physics department is an exception.

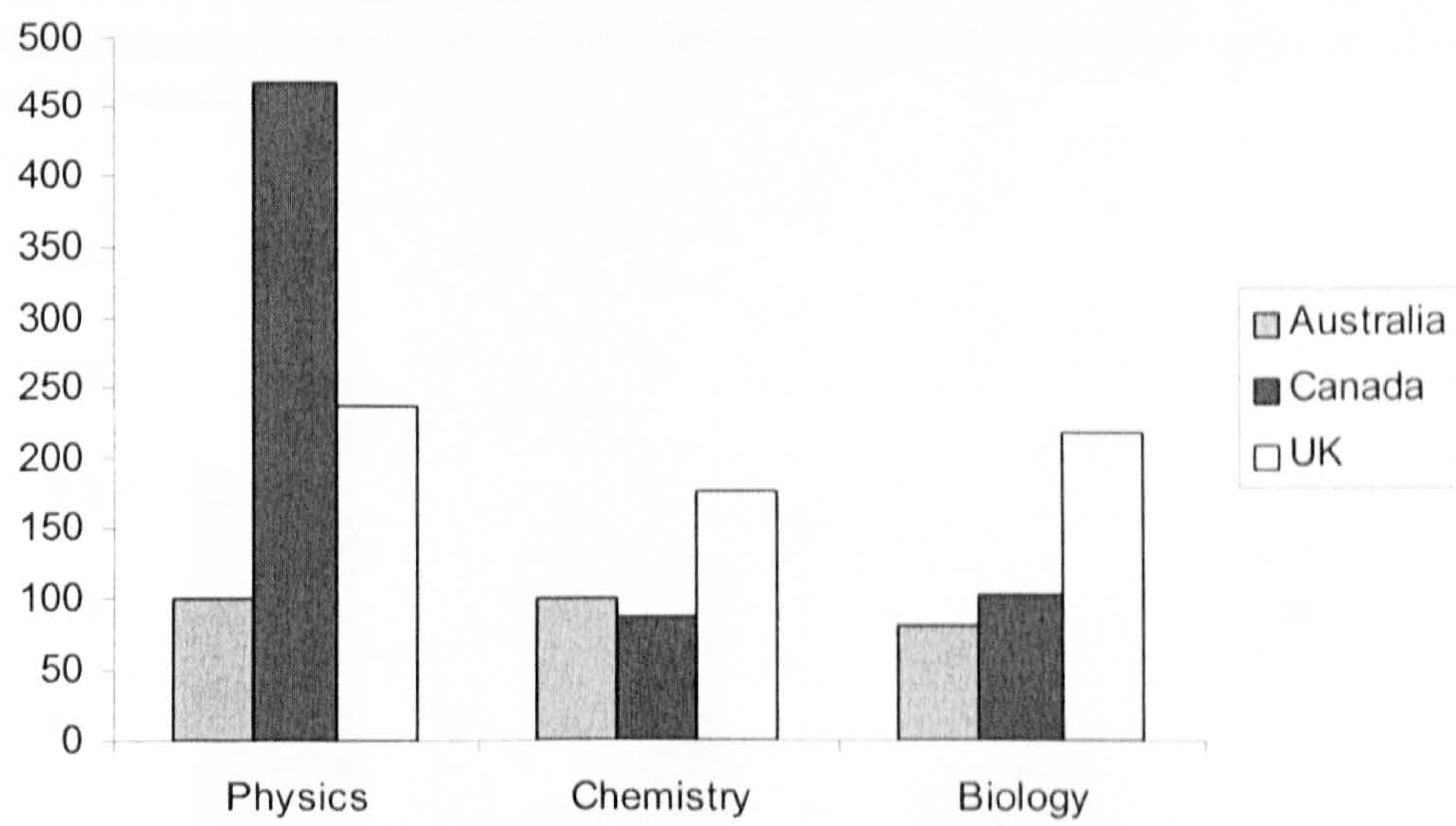


Figure 7.3 Median number of external inlinks from the whole Web (AltaVista data)

4. Proportions of national inlinks. As shown in figure 7.4, the proportions of national inlinks for the chemistry and biology departments in Australia, and physics departments in the UK are the largest. These three sets of departments are the most nationally visible in each of the three disciplines. The Australian physics, Canadian biology and UK chemistry departments' proportion of national inlinks are the smallest amongst the three countries, in other words the sets of departments are the most internationally visible.

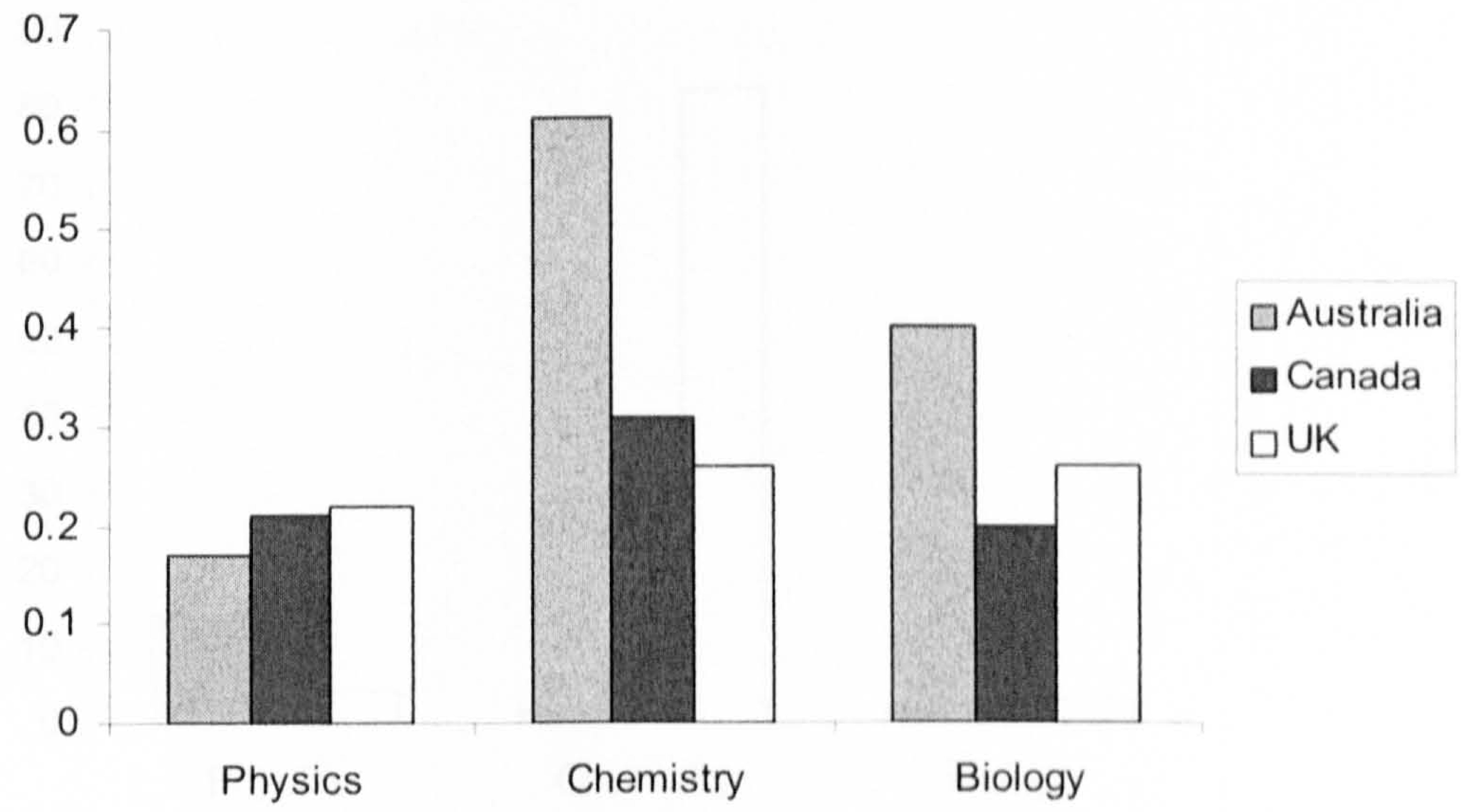


Figure 7.4 Proportion of national inlinks by country (AltaVista data)

7.2.1.2 National Peer Interlinks

1. Interconnection rates. As illustrated in figure 7.5, the Canadian physics and biology interconnected the most amongst the three countries, while the UK chemistry departments interconnected the most. The UK physics and biology departments interconnected the fewest compared with their peers in Canada and Australia.

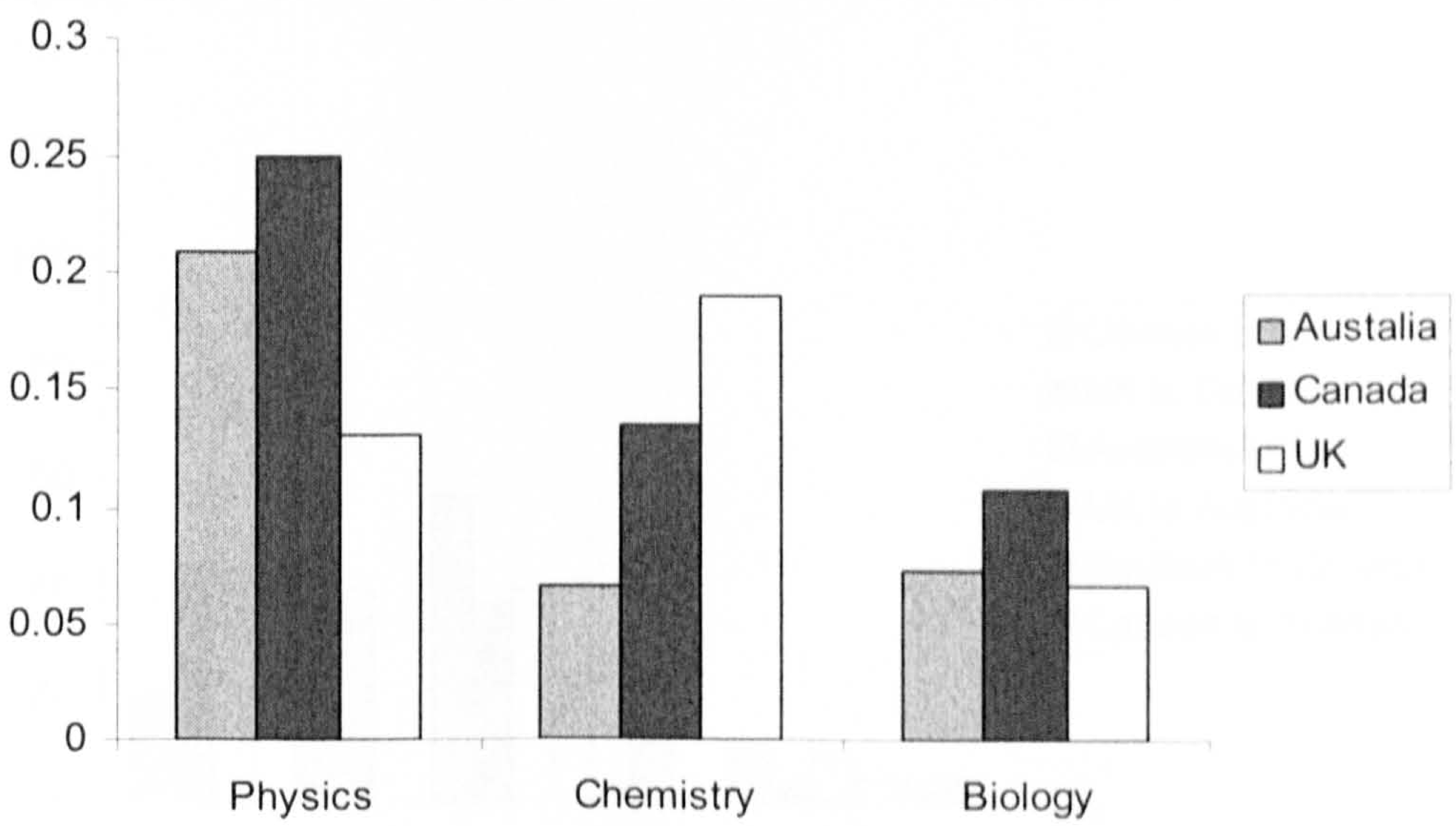


Figure 7.5 Interconnection rates by country (SocSciBot3 data)

2. Median number of national peer inlinks. As shown in figure 7.6, the UK chemistry departments attract the largest median inlinks from its peers, while the UK physics departments attract the smallest median peer inlinks. Canadian physics and biology departments attract the largest median inlinks from their peers amongst the three countries.

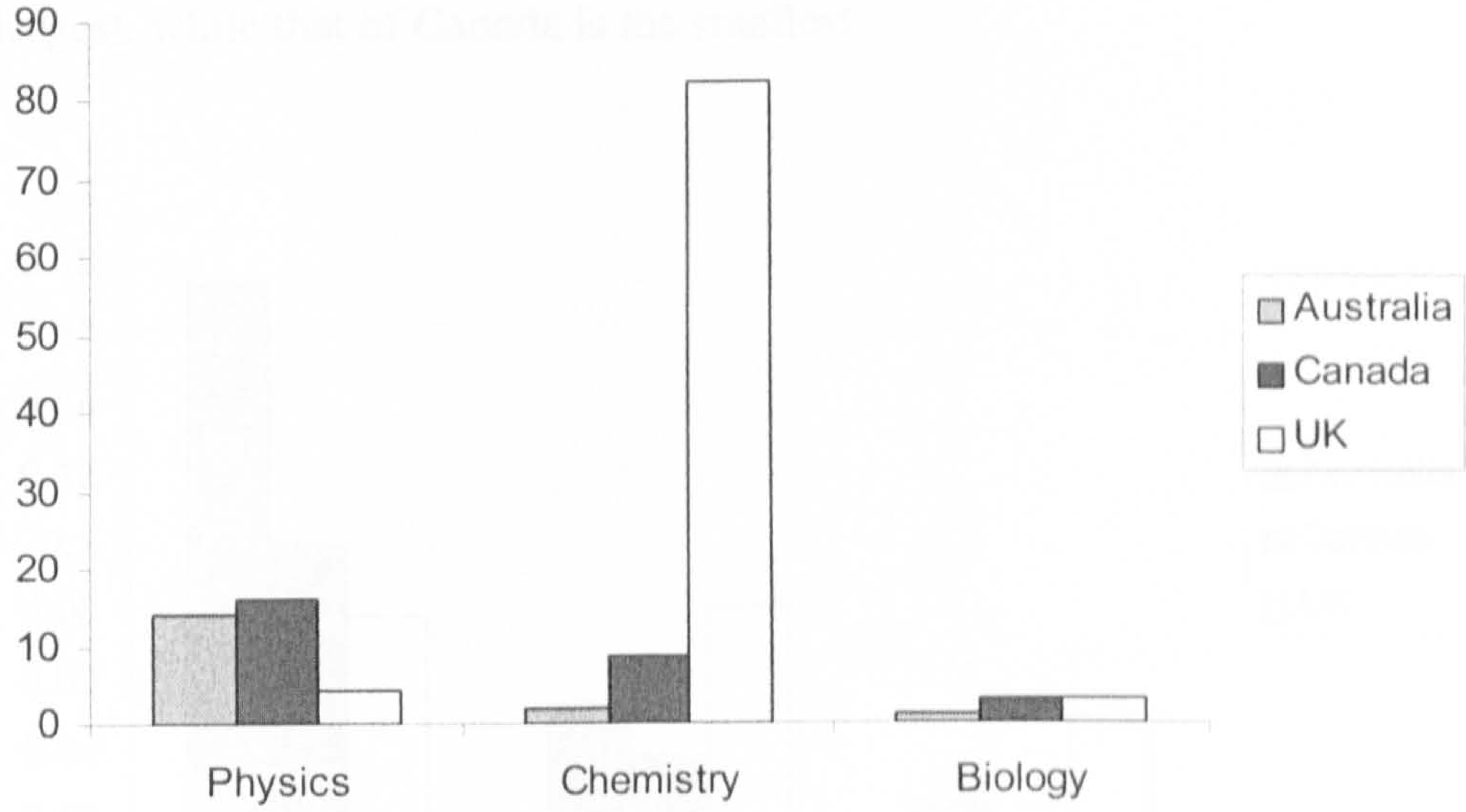


Figure 7.6 Median inlinks from peer departments by country (SocSciBot3 data)

7.2.1.3 International Peer Interlinks

1. Link Propensities. Figure 7.7 illustrates the Link Propensities between each pair of countries' departments, which are from the same discipline. For physics, the link propensity from the UK to Australia is the largest, while that from Canada to the UK is the smallest. For chemistry, the link propensity from Canada to the UK is the largest, while that from Australia to Canada is the smallest. For biology, the link propensity from Canada to the UK is the largest, while that from UK to Canada is the smallest.

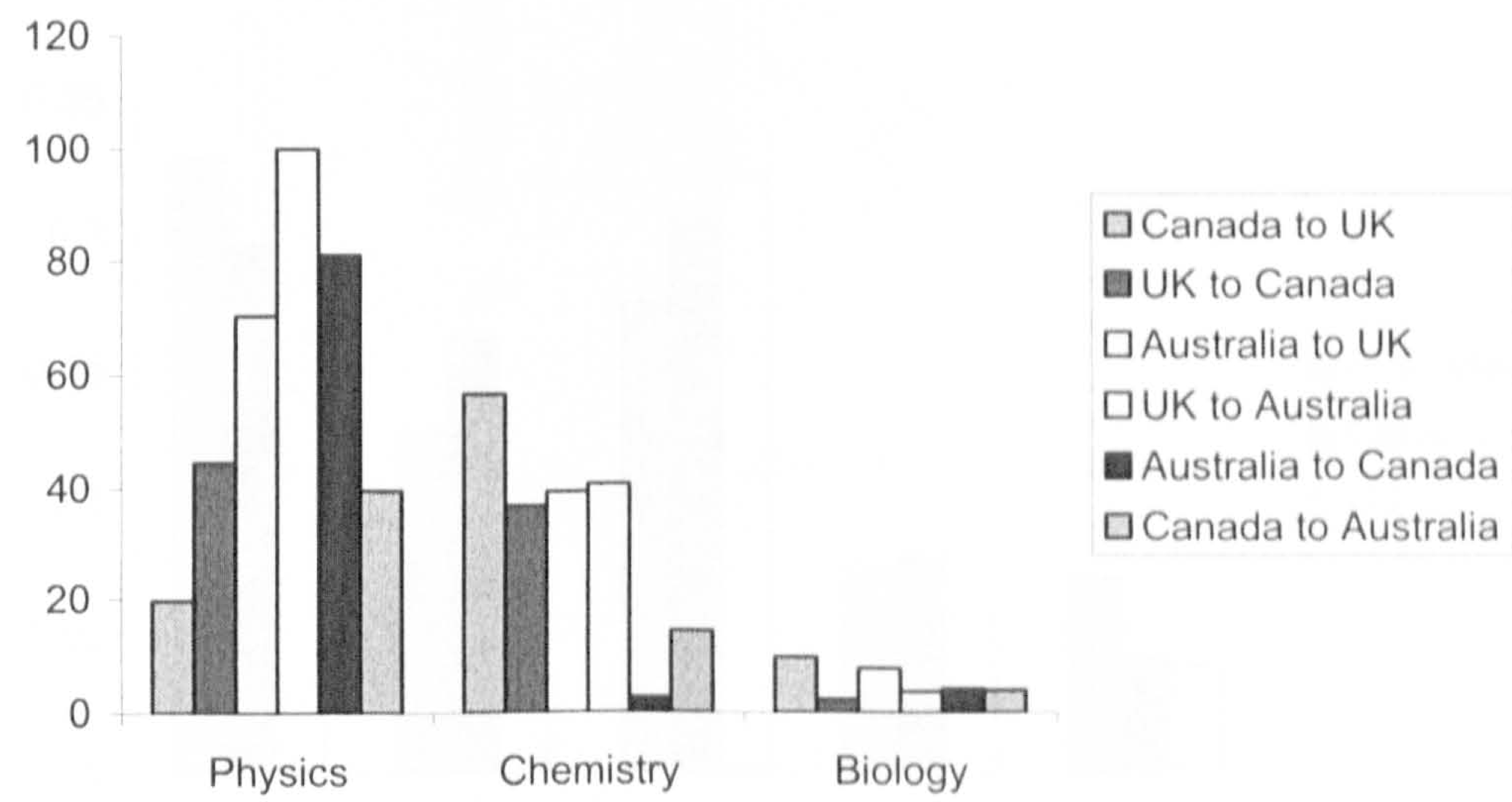


Figure 7.7 Link propensities (AltaVista data)

2. Adapted mean international peer inlinks. Figure 7,8 illustrates the adapted mean international peer inlinks. Amongst the three countries, the Australian physics departments' adapted mean value of international peer inlinks is the largest, while that of the UK is the smallest. The UK chemistry departments' mean international peer inlinks is the largest, while that of Canada is the

smallest. The UK biology departments' mean international peer inlinks is the largest, while that of Canada is the smallest.

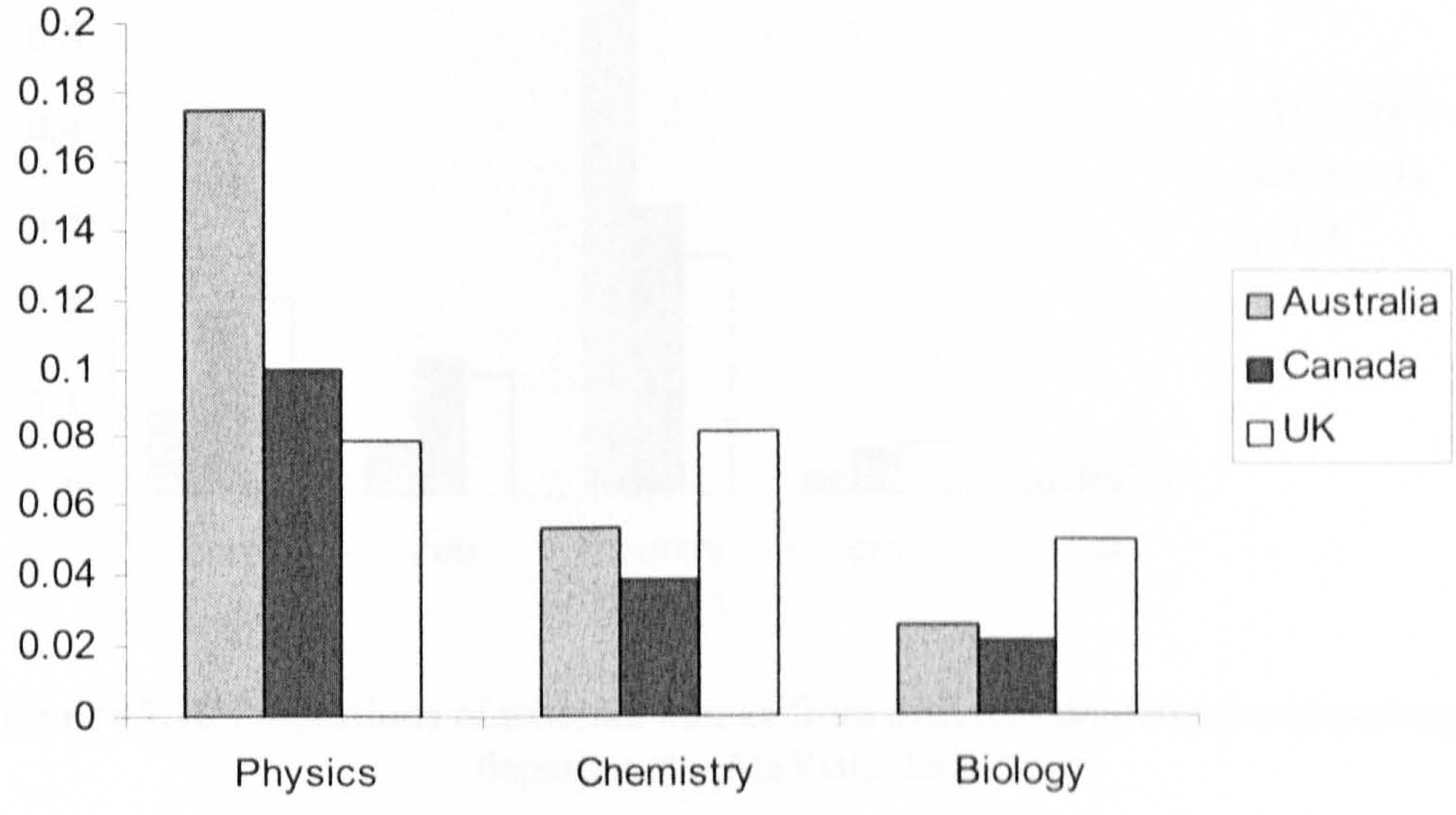


Figure 7.8 Mean international peer inlinks (AltaVista data)

7.2.1.4 Interactions with Different Top Level Domains

1. Proportions of external inlinks from different domains. Figures 7.9 to 7.11 illustrate the proportions of external inlinks from different domains for each set of departments along county lines. Canadian physics and chemistry departments attract the largest proportions of edu inlinks amongst the three countries.

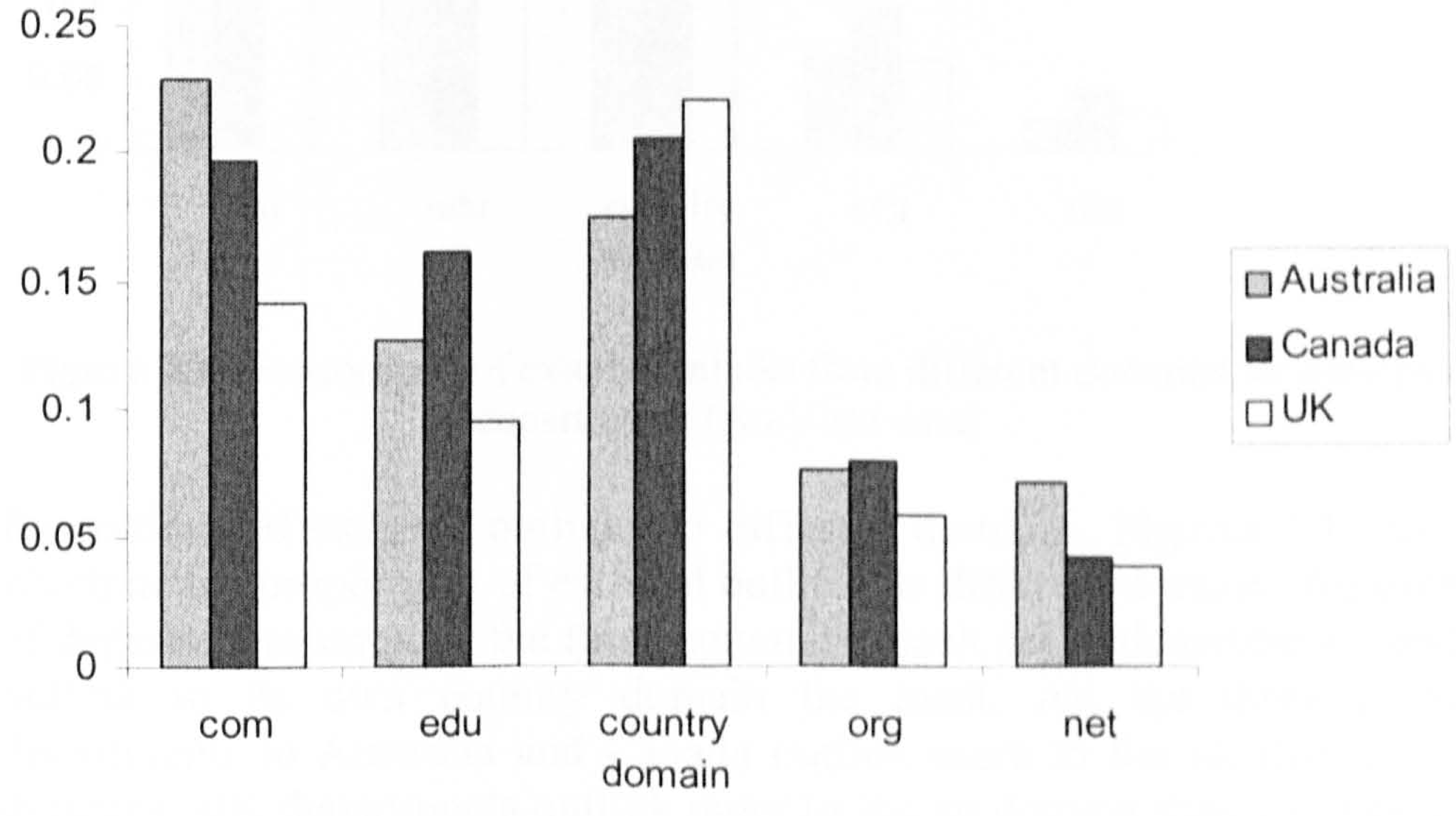


Figure 7.9 Proportions of external inlinks from different domains for physics departments (AtaVista data)

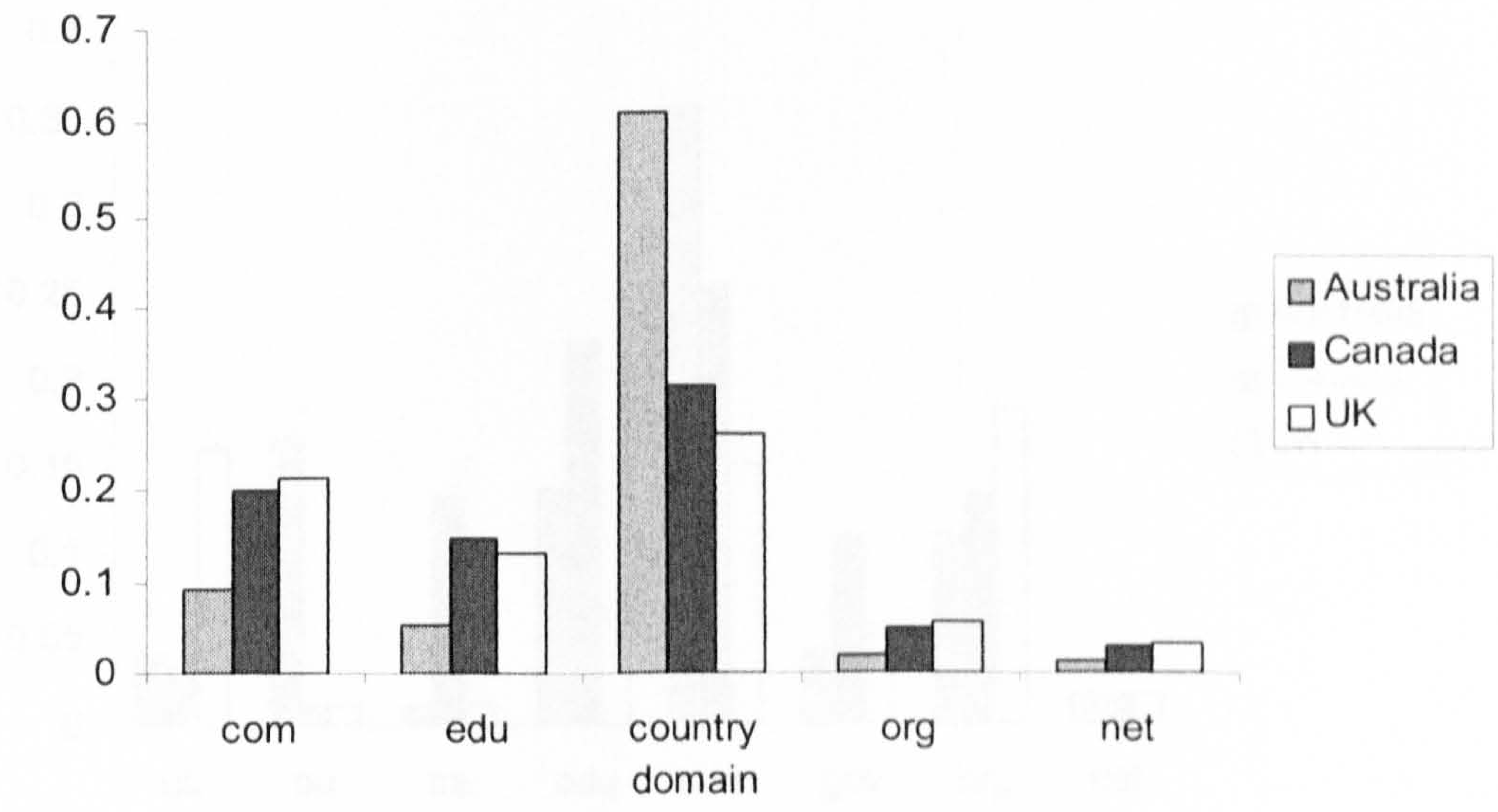


Figure 7.10 Proportions of external inlinks from different domains for chemistry departments (AtaVista data)

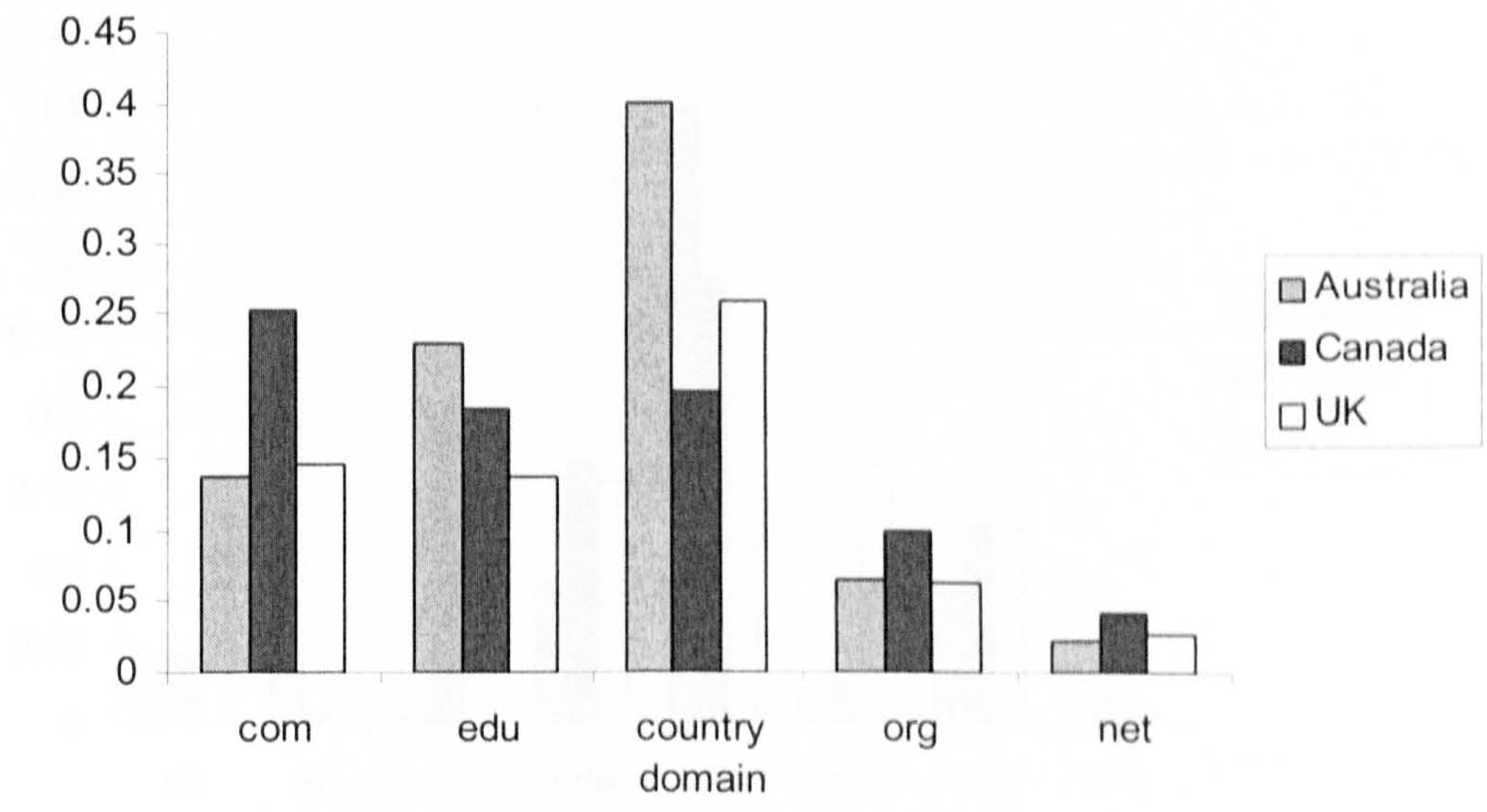


Figure 7.11 Proportions of external inlinks from different domains for biology departments (AtaVista data)

2. Proportions of external outlinks to different domains. Figures 7.12 to 7.14 illustrate the proportions of external outlinks to different domains for each set of departments amongst the three countries. Each set of departments tends to outlink to its own country domain the most. All the three types of departments in Australia and Canada outlink more to the uk than ca or au domains. UK departments outlink more to the au domain than ca. Australian departments outlink to the com domain the most, those from the UK the least. Canadian departments outlink to the edu domain the most amongst the three countries, while the Australian physics and chemistry departments outlink to the edu domain the least. Australian departments' proportions of outlinks to com domain are the largest amongst the three countries, while those for the UK departments are the smallest.

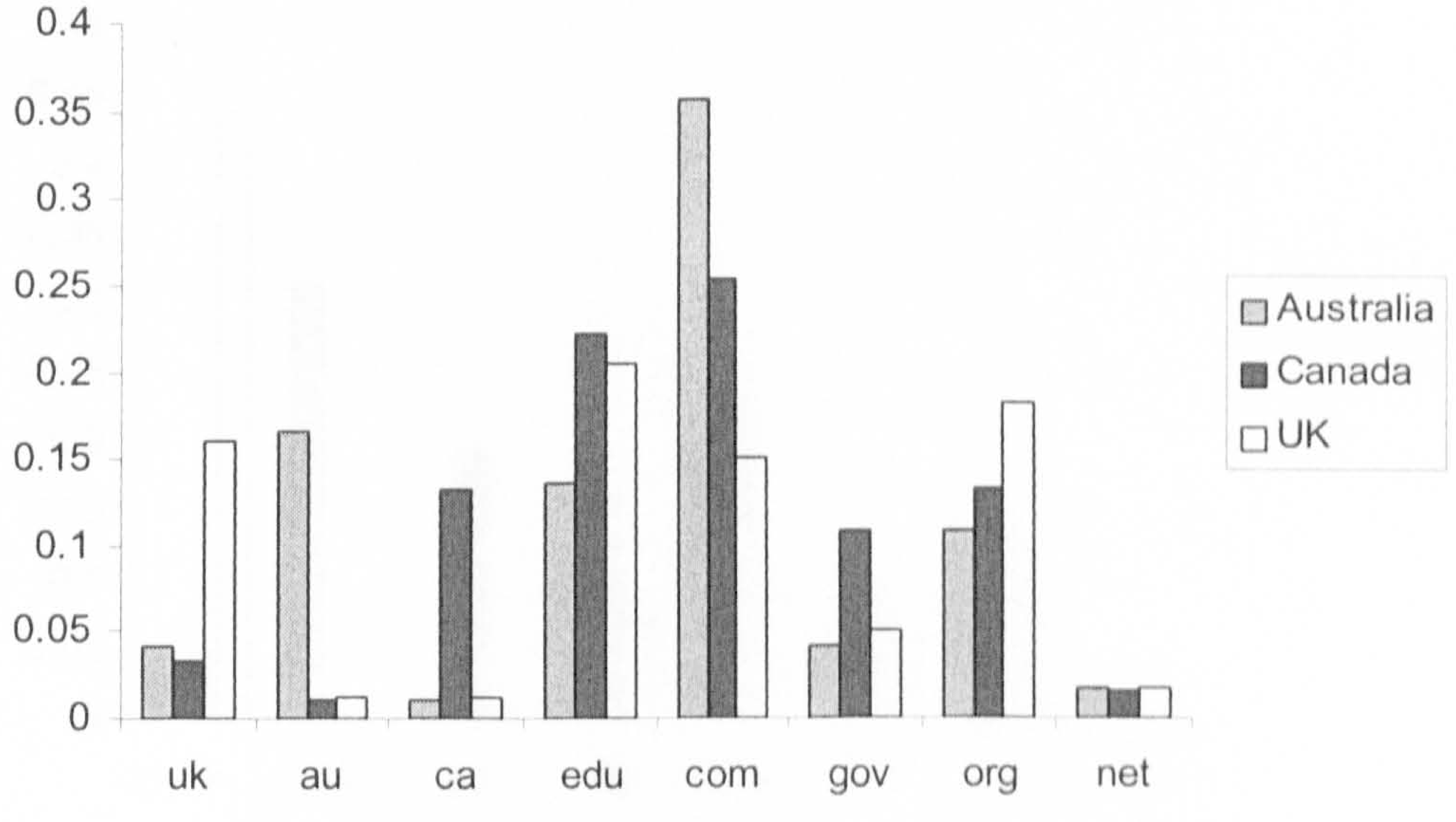


Figure 7.12 Proportions of outlinks to different domains for the physics departments (SocSciBot data)

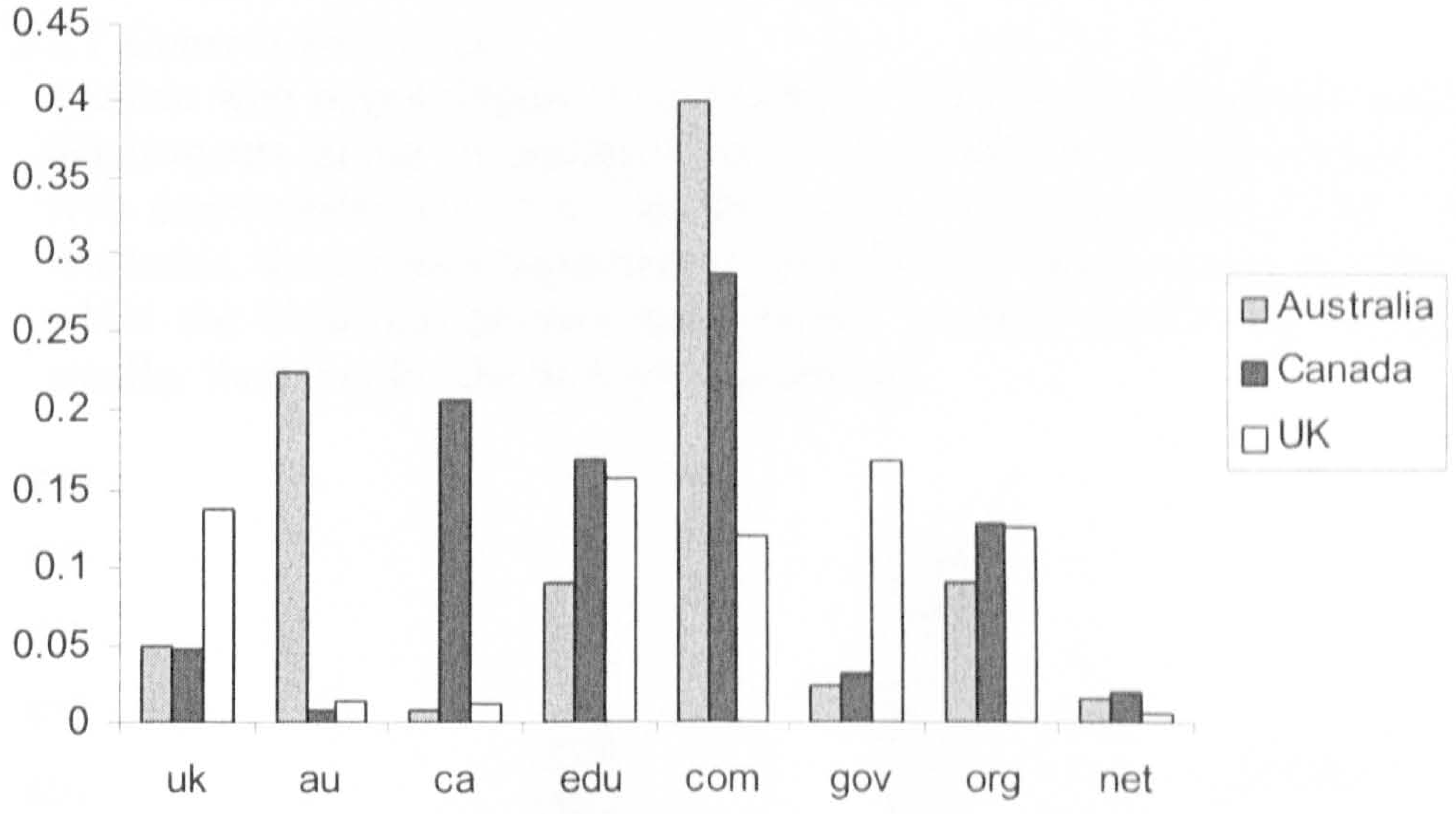


Figure 7.13 Proportions of outlinks to different domains for the chemistry departments (SocSciBot3 data)

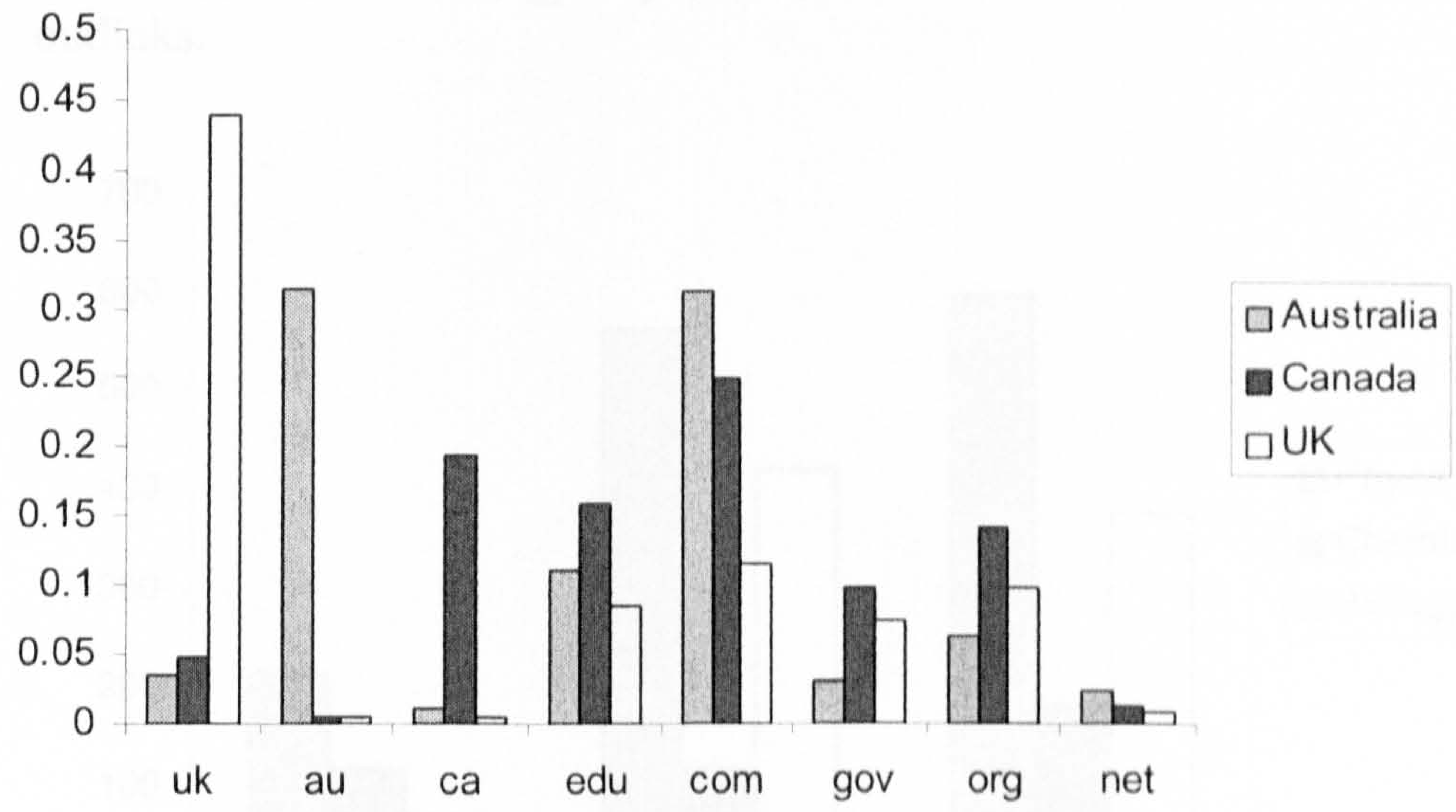


Figure 7.14 Proportions of links to different domains for the biology departments (SocSciBot3 data)

7.2.2 Disciplinary Differences

7.2.2.1 General Web Uses

1. Median web pages. Figure 7.15 illustrates median web pages for each set of departments along disciplinary lines. The chemistry departments' median web page counts are all the smallest in the three countries. In the UK and Australia, the physics departments' median web page counts are the largest, while the Canadian physics departments' median number of web pages is smaller than that for the biology departments.

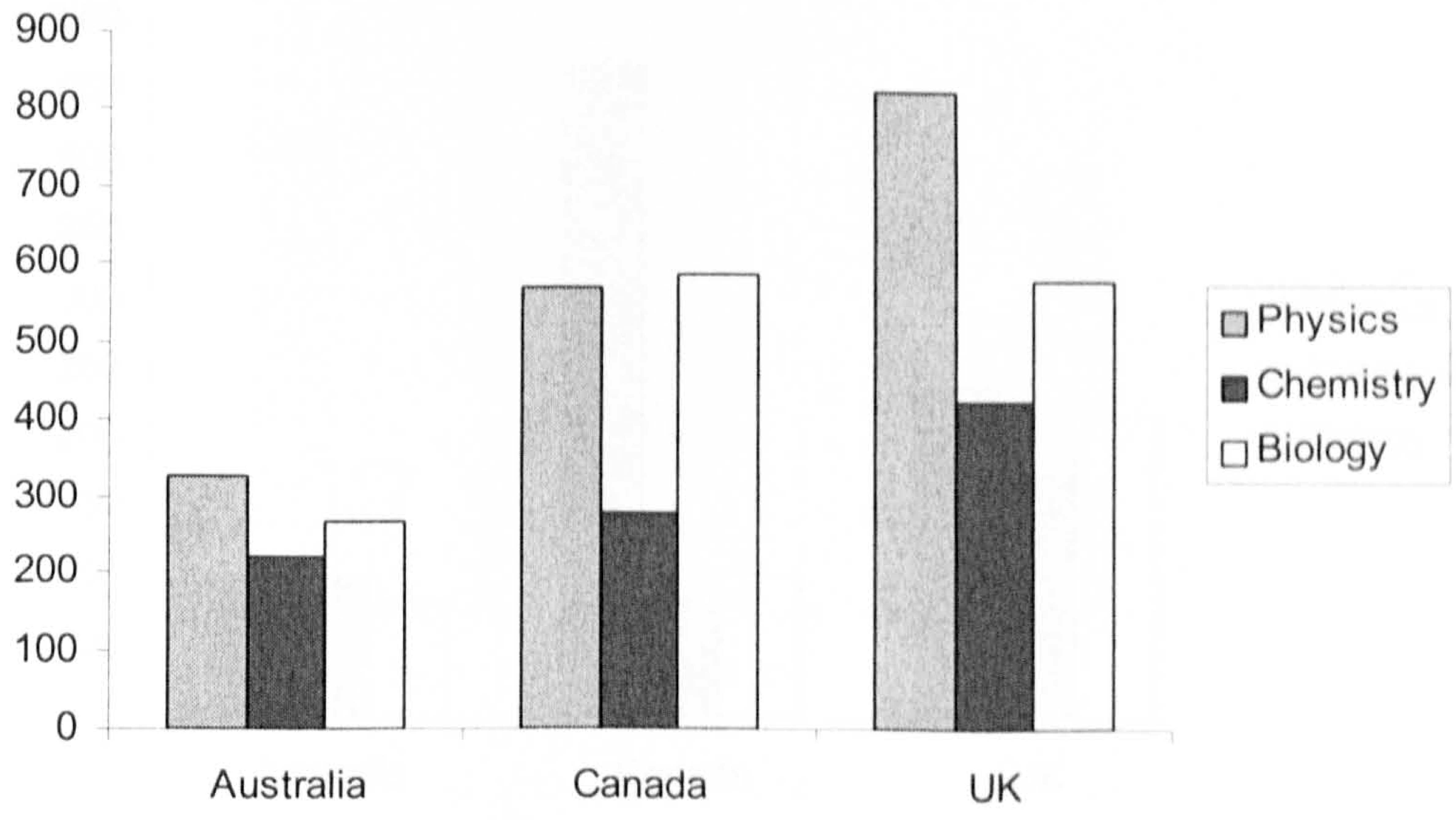


Figure 7.15 Median web pages by discipline (SocSciBot3 data)

2. Median external outlinks. Figure 7.16 illustrates median external outlinks for each set of departments along disciplinary lines. The physics departments' median external outlinks again are the largest of all the three countries, while those for the chemistry departments in the UK and Canada are the smallest.

For Australia, the biology departments have the smallest median external outlinks.

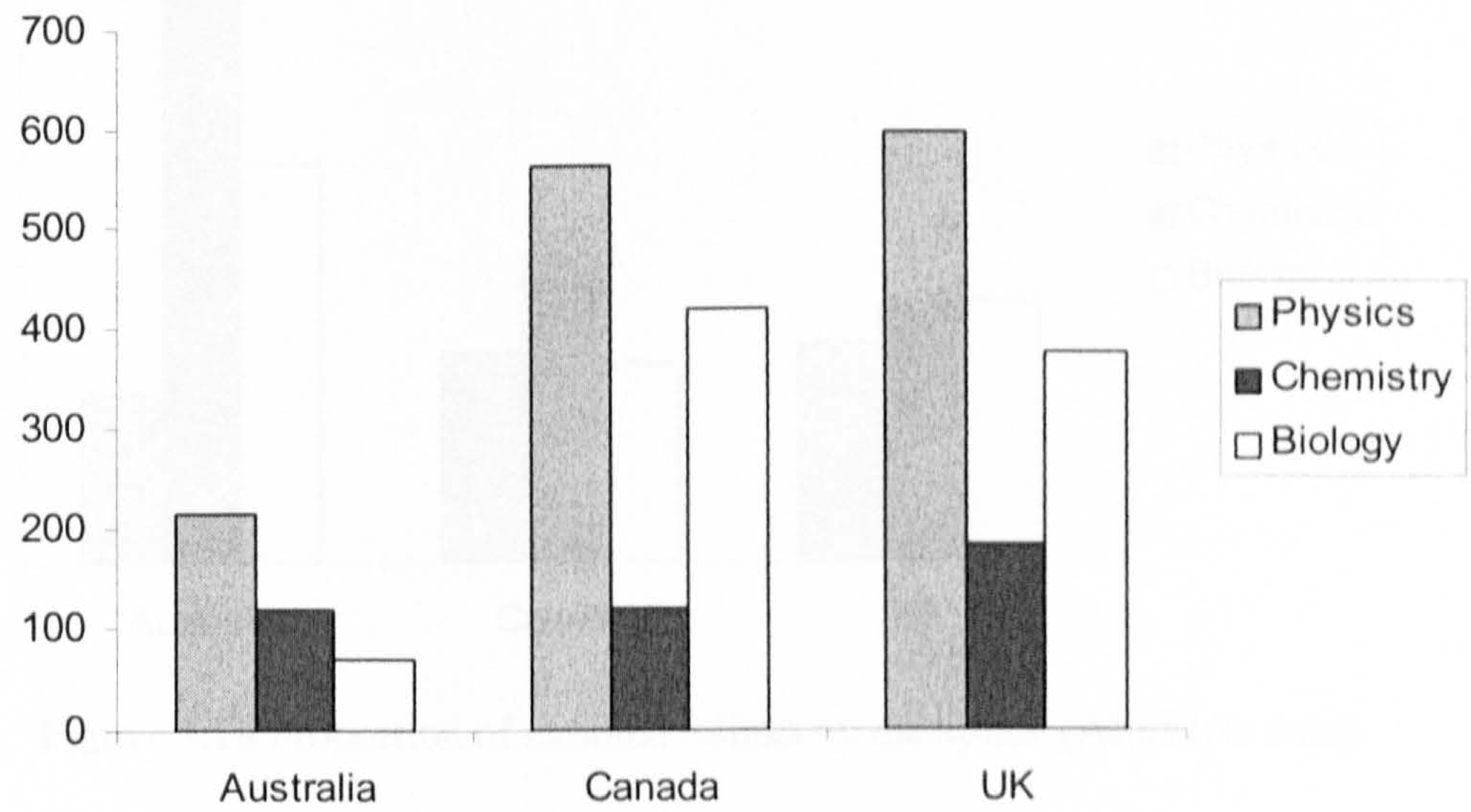


Figure 7.16 Median external outlinks by discipline (SocSciBot3 data)

3. Median number of external inlinks from the whole Web. Figure 7.17 illustrates each set of departments' median number of inlinks from the whole Web along disciplinary lines. Generally, physics departments attract the largest number of inlinks from the whole Web compare with other disciplines, while chemistry departments attract the smallest with exceptions.

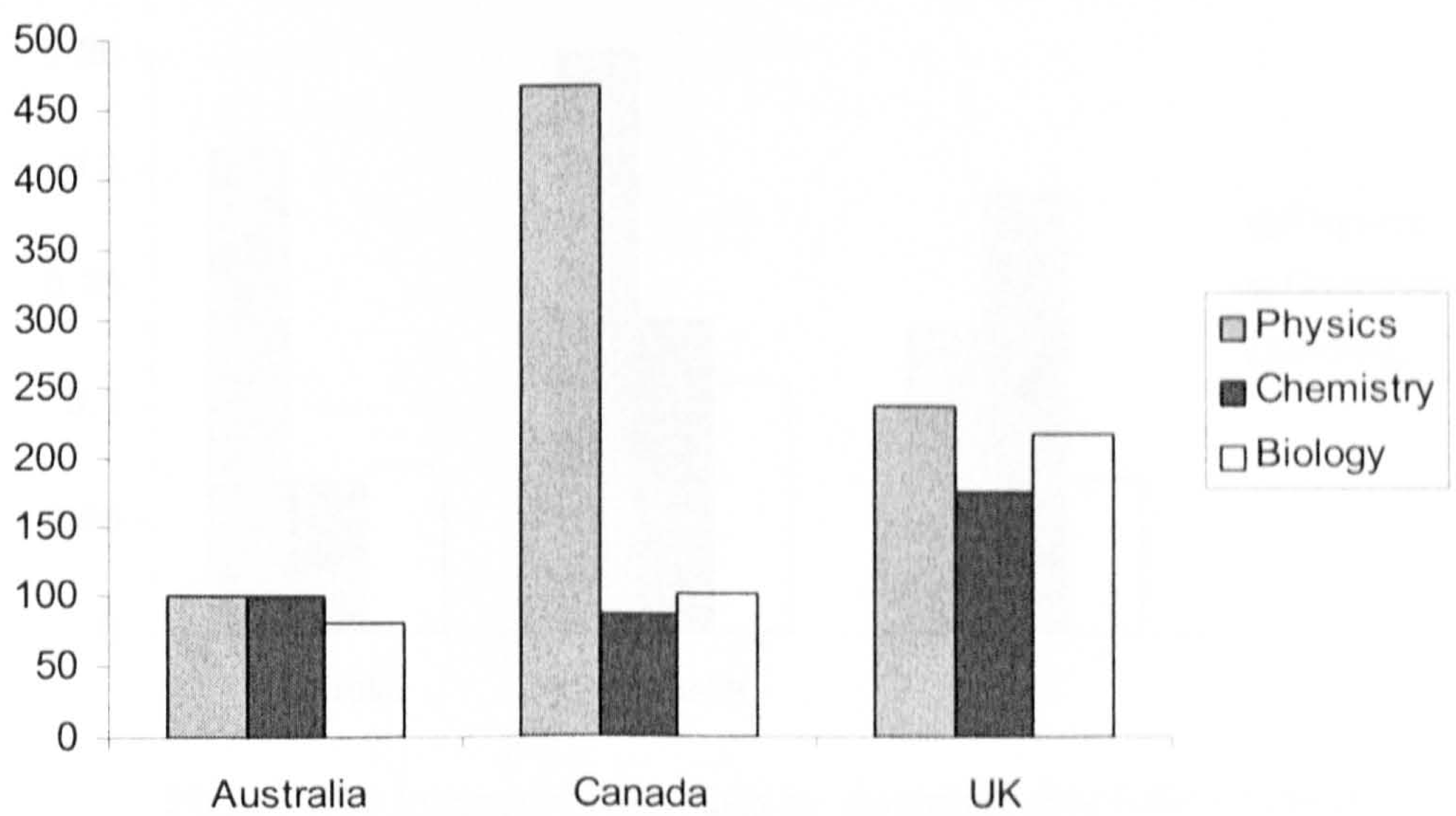


Figure 7.17 Median external inlinks from the whole Web by discipline (AltaVista data)

4. Proportions of national inlinks. Figure 7.18 shows the proportion of national inlinks for each set of departments along disciplinary lines. The chemistry departments' proportion of national inlinks is the largest in the three countries. Both in the UK and Australia, the physics departments' proportion of national inlinks is the smallest.

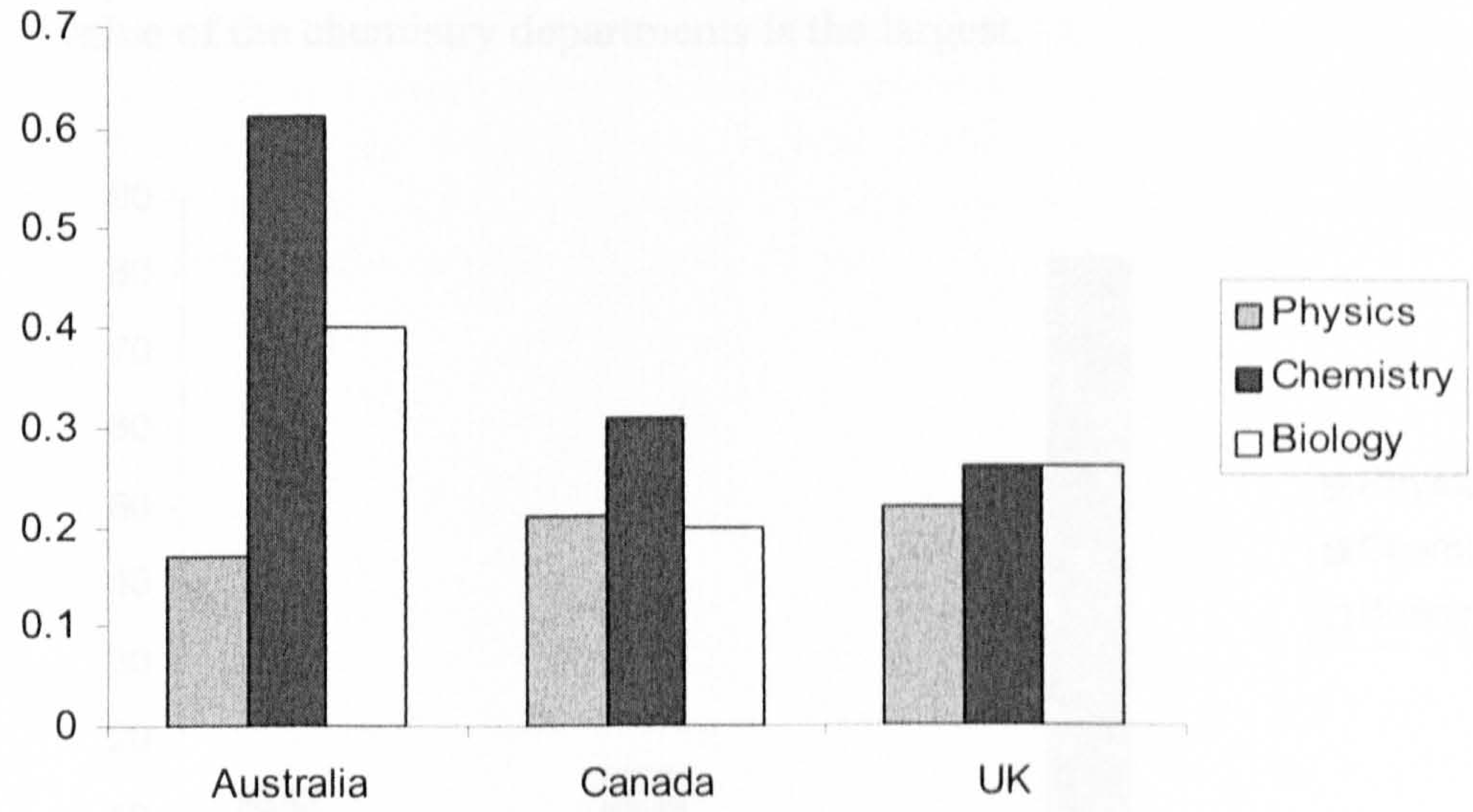


Figure 7.18 Proportion of national inlinks by discipline (AltaVista data)

7.2.2.2 National Peer Interlinks

1. Interconnection rates. Figure 7.19 illustrates the interconnection rate of each set of departments along disciplinary lines. In Australia and Canada, physics departments interconnect the best, while in the UK, the chemistry departments interconnect the best. In the UK and Canada, the biology departments' interconnection rate is the smallest, while in Australia, the chemistry departments' interconnection rate is the smallest.

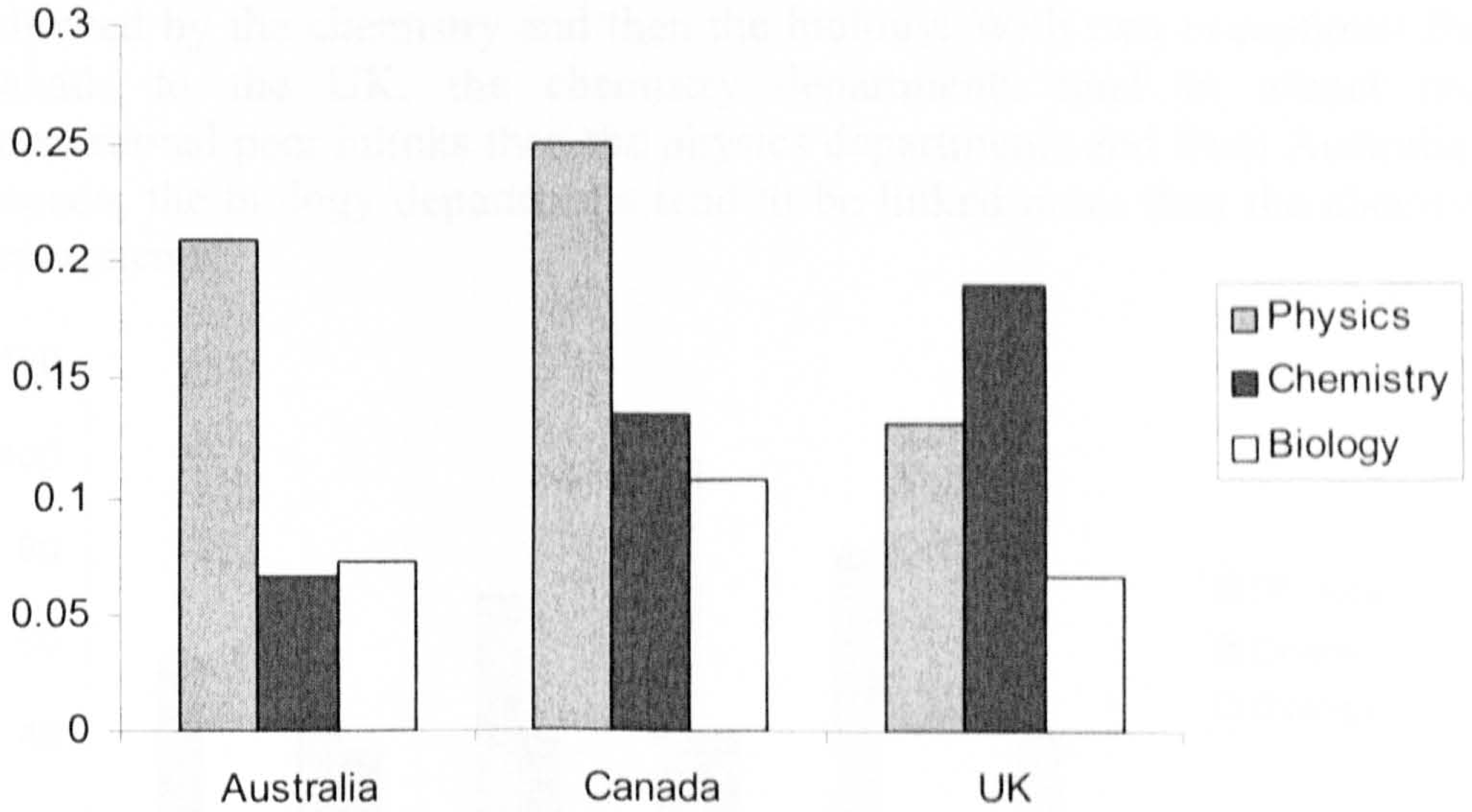


Figure 7.19 Interconnection rates by discipline (SocSciBot3 data)

2. Median national peer inlinks. Figure 7.20 illustrates the median number of national peer inlinks for each set of departments along disciplinary lines. The median number of peer inlinks for the UK chemistry departments is extremely large, dwarfing other sets of departments' values. Both in Canada and Australia, the physics departments' median peer inlinks are the largest, those for the biology departments are the smallest. In the UK, the biology

departments' median number of peer inlinks is the smallest; however, the value of the chemistry departments is the largest.

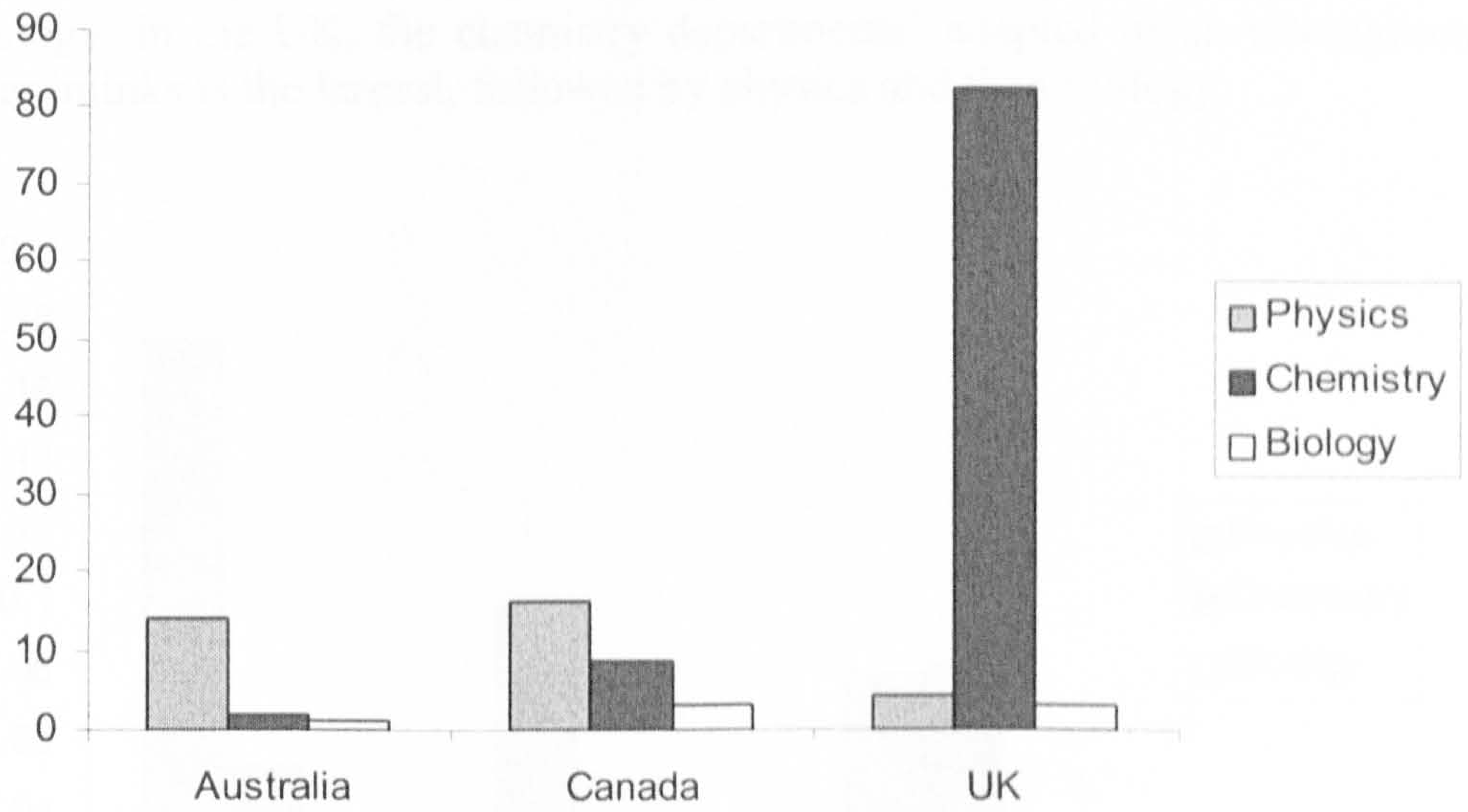


Figure 7.20 Median peer inlinks in the same country by discipline (SocSciBot3 data)

7.2.2.3 International Peer Interlinks

1. Link Propensities. Figure 7.21 illustrates the Link Propensity for each set of departments along disciplinary lines. The physics departments in each country tend to be linked the most from their peers in other countries, followed by the chemistry and then the biology. With two exceptions: from Canada to the UK, the chemistry departments tend to attract more international peer inlinks than the physics departments and from Australia to Canada, the biology departments tend to be linked more than the chemistry departments.

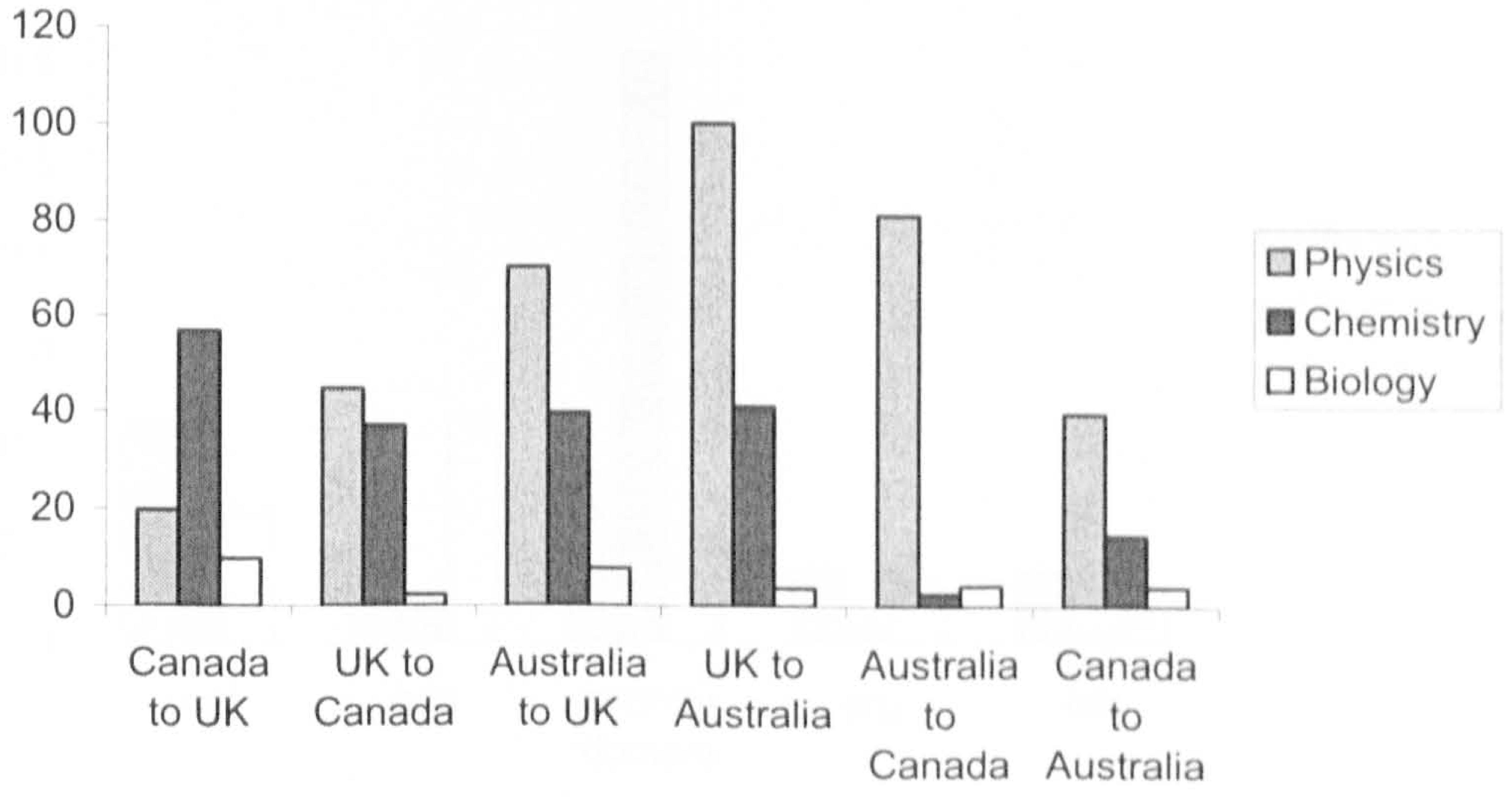


Figure 7.21 Link propensity (with denominator the product of the numbers of staff members) by discipline (AltaVista data)

2. Adapted mean international peer inlinks. Figure 7.22 illustrates the adapted mean international peer inlinks for each set of departments along disciplinary lines. Both in Australia and Canada, the physics departments' adapted mean international peer inlinks is the largest, followed by chemistry and then biology. In the UK, the chemistry departments' adapted mean international peer inlinks is the largest, followed by physics and then biology.

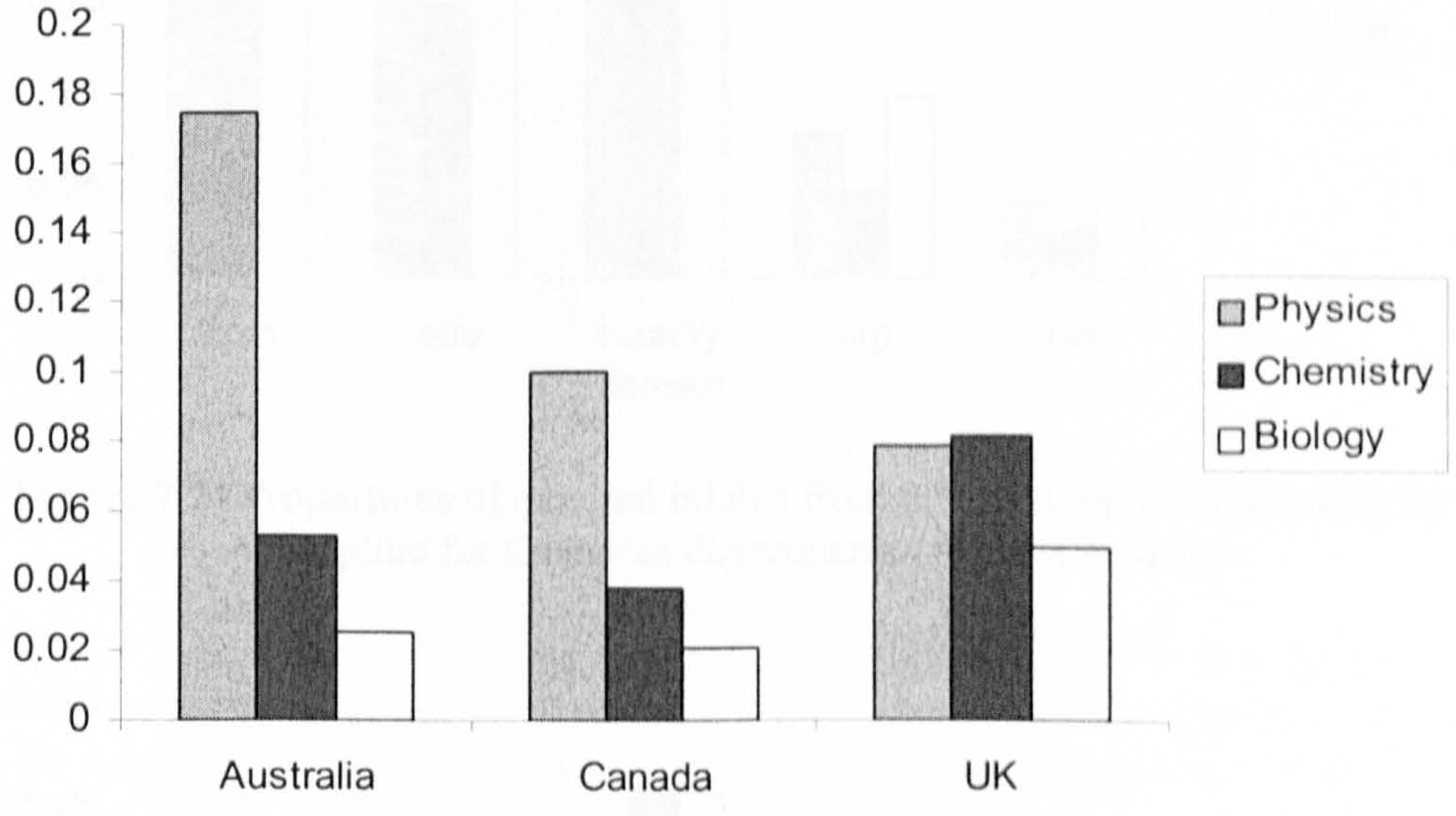


Figure 7.22 Adapted mean international peer inlinks by discipline (AltaVista data)

7.2.2.4 Interactions with Different Top Level Domains

1. Proportions of external inlinks from different top level domains. Figure 7.23 to 7.25 illustrate the proportion of external inlinks for each set of departments along disciplinary lines.

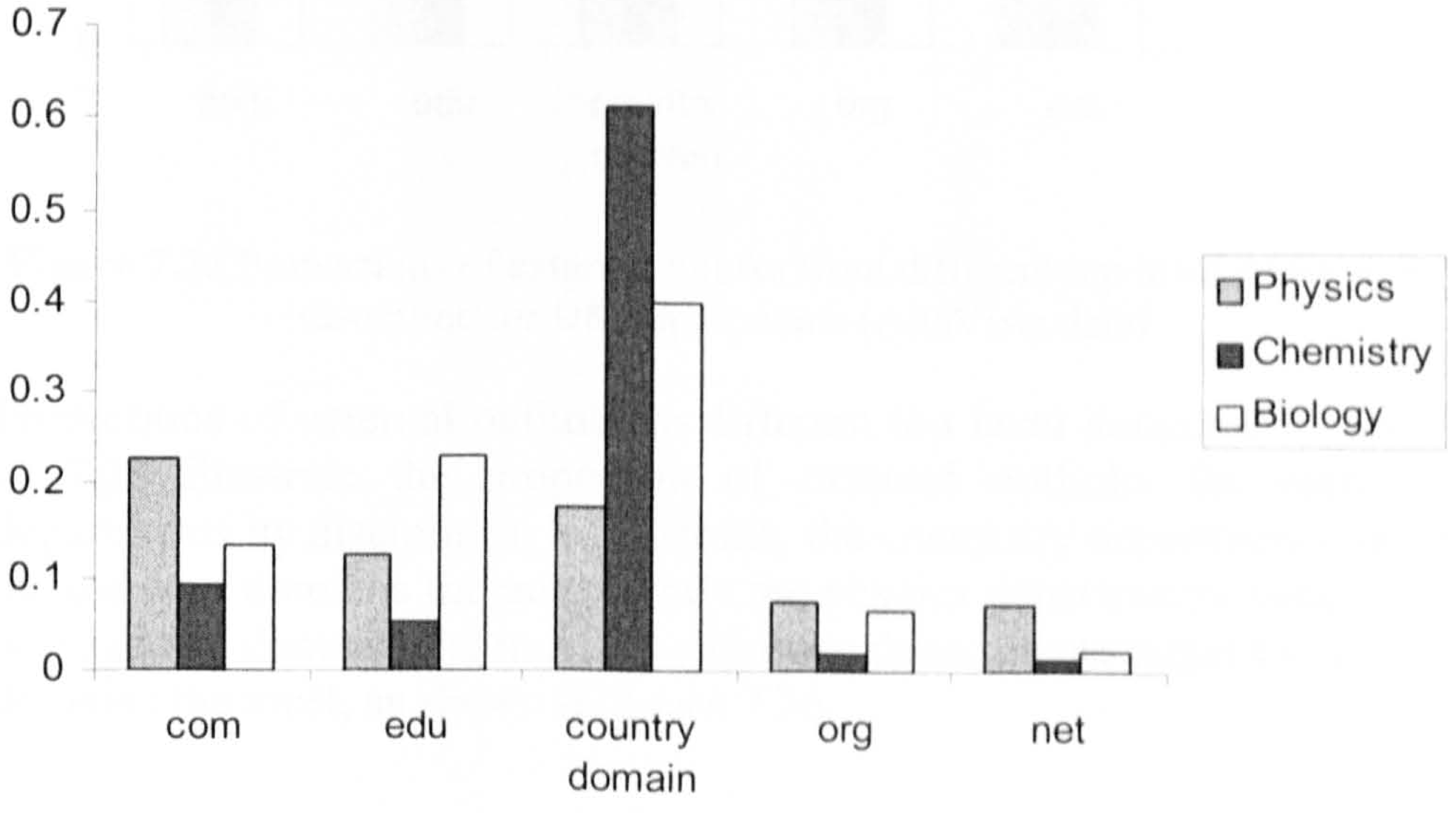


Figure 7.23 Proportions of external inlinks from different top level domains by discipline for Australian departments (AltaVista data)

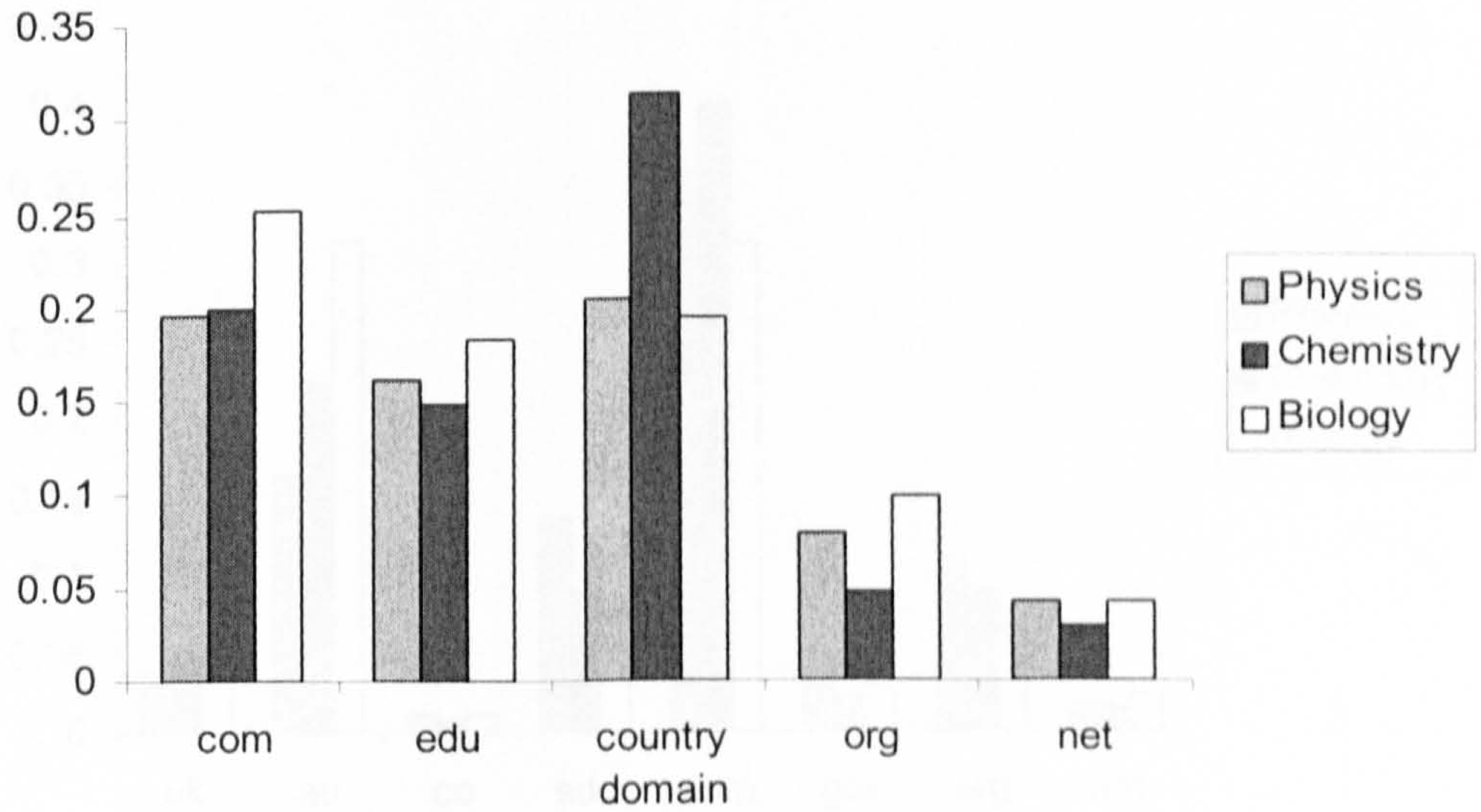


Figure 7.24 Proportions of external inlinks from different top level domains by discipline for Canadian departments (AltaVista data)

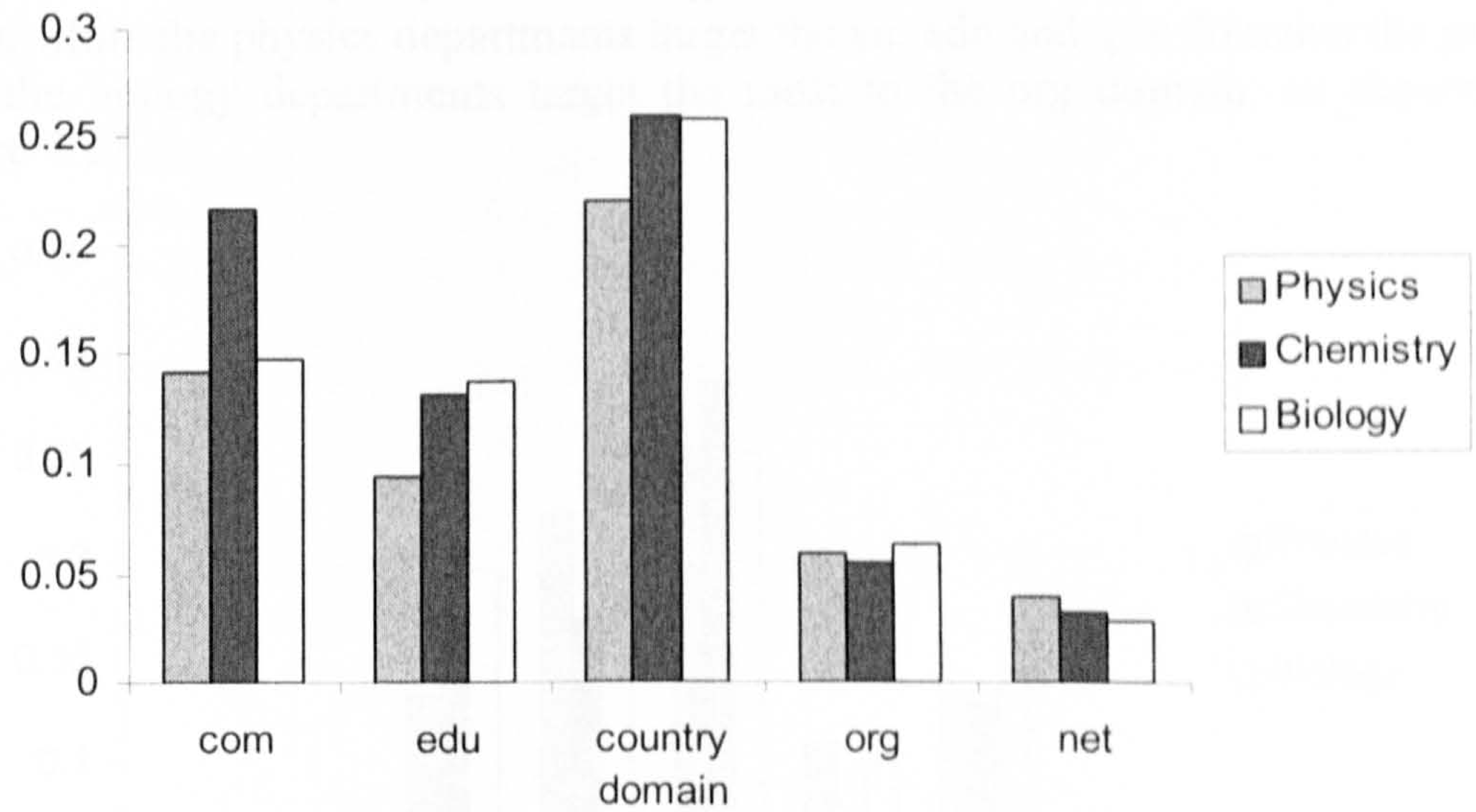


Figure 7.25 Proportions of external inlinks from different top level domains by discipline for UK departments (AltaVista data)

2. Proportions of external outlinks to different top level domains. Figures 7.26 to 7.28 illustrate the proportion of external outlinks for each set of departments by disciplines. In Australia, the chemistry departments target the uk and com domains the most, while the physics departments target the edu, gov and org domains the most. The biology departments target the au and net domains the most, as shown in figure 7.26.

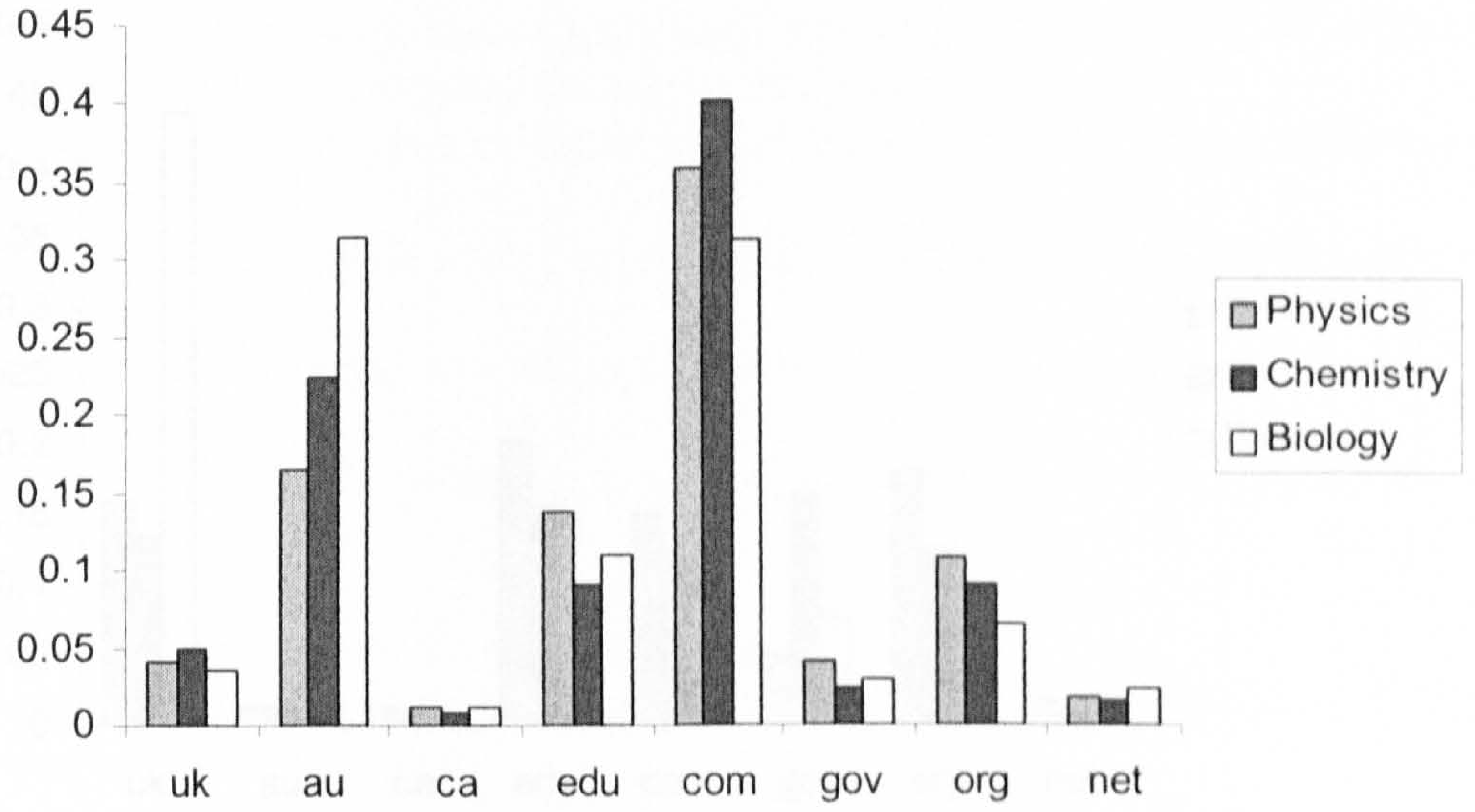


Figure 7.26 Proportions of external outlinks to different top level domains for the Australian departments by discipline (SocSciBot3 data)

In Canada, chemistry departments target the uk, ca, com and net domains the most, while the physics departments target the au, edu and gov domains the most and the biology departments target the most to the org domain, as shown in figure 7.27.

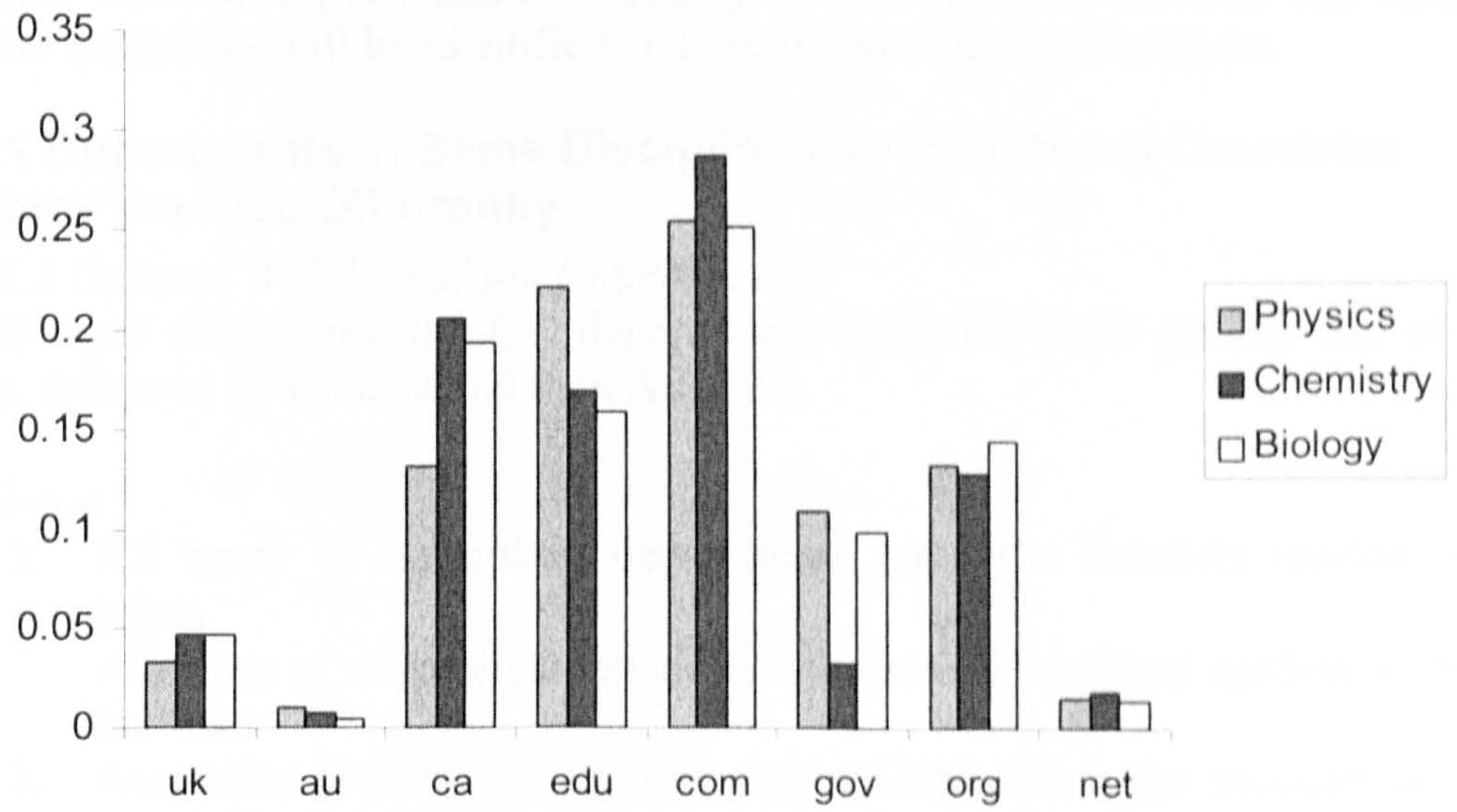


Figure 7.27 Proportions of external outlinks to different top level domains for the Canadian departments by discipline (SocSciBot3 data)

In the UK, biology departments target the uk domain the most; the physics departments target the ca, edu, com, org and net domains the most; and the chemistry departments target the au and gov domains the most, as shown in figure 7.28.

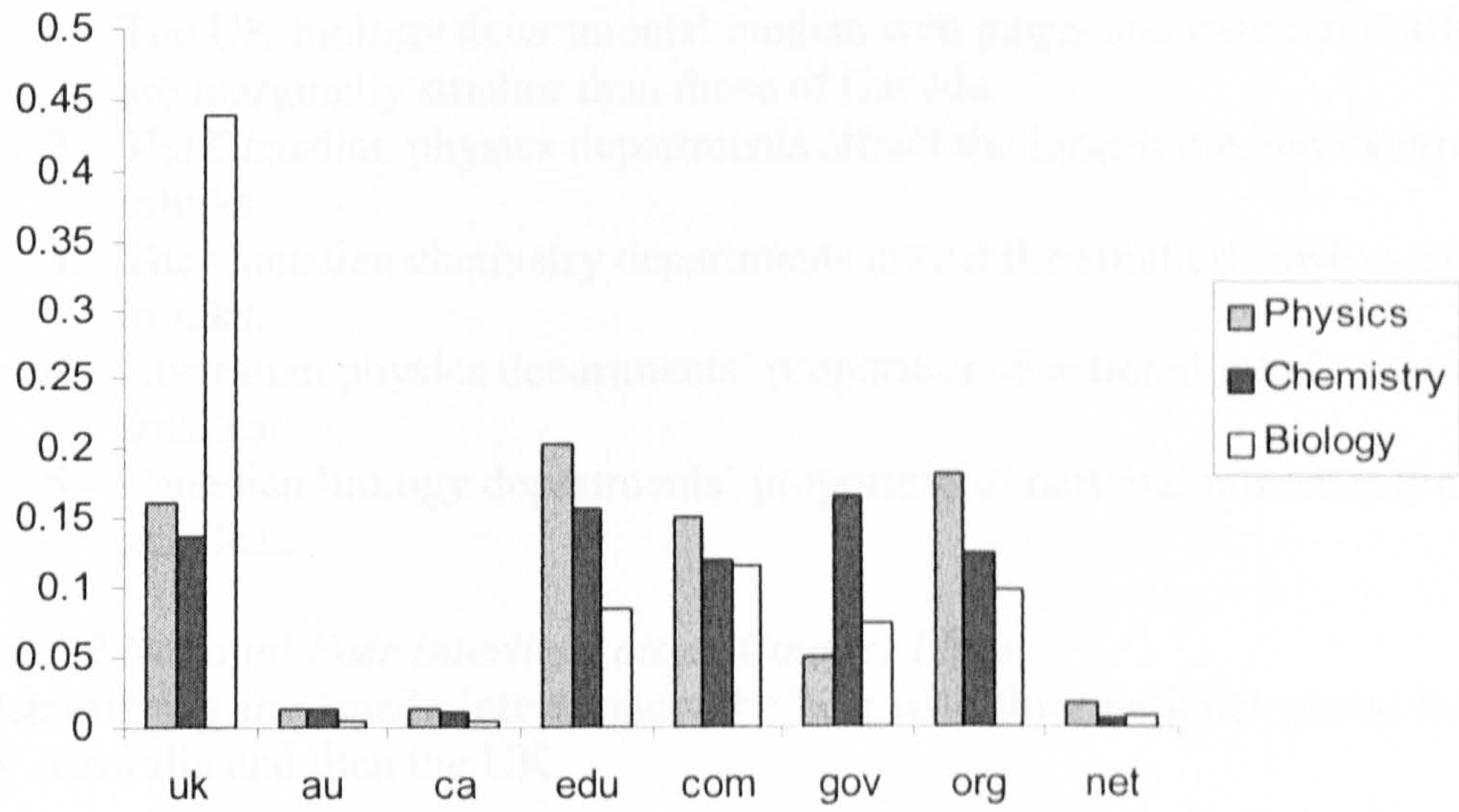


Figure 7.28 Proportions of external outlinks to different top level domains for the UK departments by discipline (SocSciBot3 data)

7.3 Summary

The link patterns along country and disciplinary lines are difficult to generalize, partly because of the large number of results. Patterns are listed with regard to the four different aspects and followed by evidences and anomalies. The reasons for the anomalies will be identified and discussed in the next chapter.

7.3.1 Departments in Same Disciplines from Different Countries Employ the Web Differently

7.3.1.1 General Web Use along Country Lines

In all three disciplines, the UK departments make the most general use of the Web, followed by Canada and then Australia.

Evidence

1. All types of Australian departments have the smallest median web pages.
2. All types of Australian departments create the smallest median external outlinks.
3. Australian physics and biology departments attract the smallest median external inlinks.
4. Australian chemistry and biology departments' proportion of national inlinks are the largest.
5. UK physics and chemistry departments have the largest median web pages.
6. UK physics and chemistry departments create the largest median external outlinks.
7. UK chemistry and biology departments attract the largest median external inlinks.
8. UK chemistry departments' proportion of national inlinks is the smallest.

Anomalies

1. The UK biology departments' median web pages and external outlinks are marginally smaller than those of Canada.
2. The Canadian physics departments attract the largest median external inlinks.
3. The Canadian chemistry departments attract the smallest median external inlinks.
4. Australian physics departments' proportion of national inlinks is the smallest.
5. Canadian biology departments' proportion of national inlinks is the smallest.

7.3.1.2 National Peer Interlinks along Country Lines

Departments in Canada interconnect the best with their national peers, followed by Australia and then the UK.

Evidence

1. Canadian physics and biology departments receive the largest median national peer inlinks.
2. Canadian physics and biology departments' have the largest interconnection rates.
3. UK physics departments' median peer inlinks and interconnection rate are the smallest
4. UK biology departments' interconnection rate is the smallest

Anomalies

1. The UK chemistry departments' median national peer inlinks and interconnection rate are the largest amongst the three countries.
2. The Australian biology departments receive smaller median national peer inlinks than those in the UK.

7.3.1.3 International Peer Interlinks Along Country Lines

Departments in the UK attract the most international peer inlinks, followed by Australia and then Canada.

Evidence

1. The UK chemistry and biology departments receive the largest adapted mean international peer inlinks, while those in Canada receive the least.
2. In chemistry, the link propensity from Canada to the UK is the largest, while that from Australia to Canada the smallest.
3. In biology, the link propensity from Canada to the UK is the largest, while that from the UK to Canada the smallest.

Anomalies

1. Australian physics departments receive the largest mean international peer inlinks, those in the UK the least.
2. In physics, the link propensity from the UK to Australia is the largest, while that from Canada to the UK the smallest.

7.3.1.4 Interactions with Different Web Areas Along Country Lines

Departments attract inlinks or create outlink differently with regard to various top level domains along country lines.

1. Canadian physics and chemistry departments attract the most edu inlinks amongst the three countries.
2. UK physics and biology departments attract the least edu inlinks amongst the three countries.
3. Australian chemistry and biology departments attract the least com inlinks amongst the three coutries.
4. With regard to the au, ca and uk domains, each set of departments outlink to its own country domains the most. Departments from Australia and Canada outlink to the uk domain the second. Departments in the UK link to the au domain the second.
5. All the three sets of Canadian departments outlink to the edu domain the most.
6. All the three sets of Australian departments outlink to the com domain the most.

7.3.2 Departments from Different Disciplines Employ the Web Differently in Each Country

7.3.2.1 General Web Uses Along Disciplinary Lines

The physics departments make the most general use of the Web, followed by biology and then chemistry.

Evidence

1. Chemistry departments have the smallest median web pages.
2. Physics departments in Australia and the UK have the largest median web pages.
3. Each set of physics departments creates the largest median external outlinks.
4. Each set of chemistry departments creates the smallest median external outlinks.
5. The physics departments attract the largest median external inlinks from the whole Web.
6. The Canadian and UK chemistry departments receive the smallest median external inlinks.
7. The chemistry departments' proportions of national inlinks are the largest in each country.
8. The Australian and UK physics departments' proportions of national inlinks are the smallest.

Anomalies

1. The Canadian physics departments' median number of web pages is marginally smaller than that of biology departments.
2. The Australian chemistry departments create larger median external outlinks than the biology departments.
3. The Australian biology departments receive the smallest median external inlinks.
4. The Canadian physics departments' national inlink proportion is marginally larger than that of the biology departments.

7.3.1.2 National Peer Interlinks Along Disciplinary Lines
Physics departments attract the most national peer inlinks, followed by Chemistry and then Biology.

Evidence

1. Australian and Canadian physics departments attract the largest median national peer inlinks
2. Australian and Canadian physics departments' interconnect rates are the smallest.
3. Biology departments receive the least median national peer inlinks.
4. Biology departments' interconnect rates are the smallest.

Anomalies

1. The UK chemistry departments' median national peer inlinks is the largest
2. The UK chemistry departments' interconnection rate is the largest.

7.3.2.3 International Peer Interlinks Along Disciplinary Lines
Physics departments attract the most international peer inlinks, followed by chemistry and then biology.

Evidence

1. Physics departments in Australia and Canada receive the largest adapted mean international peer inlinks.
2. Biology departments receive the smallest mean international peer inlinks.
3. Physics departments' link propensity is the largest in each country.
4. Biology departments' link propensity is the smallest in each country.

Anomalies

1. The UK physics departments' adapted mean international peer inlinks is smaller than that of the chemistry departments.
2. From Canada to the UK, chemistry departments' link propensity is the largest.
3. From Australia to Canada, biology departments' link propensity is larger than that of the chemistry departments.

7.3.2.4 Interactions with Different Web Areas Along Disciplinary Lines
Various types of departments attract inlinks or create outlinks to various top level domains differently.

1. Chemistry departments attract the least external inlinks from org domain in each country.
2. Australian and Canadian chemistry departments attract the least external inlinks from edu and net domains.
3. Each set of physics departments target the edu domain the most.
4. Chemistry departments in Australia and Canada target the com and uk domains the most.

7.3.3 Examples of Outstandingly Good or Bad Web Exploitation

Departments not fitting well into the patterns above are identified and listed in this section. This acts both as a warning for those having worse than expected performances on the Web relative to their research measures (see table 8.21, the world relative citation impact for each set of departments), and an indication of better web experience.

7.3.3.1 Good Web Exploitation

- Australian physics departments

Despite the fact that the set of departments has the smallest world relative citation impact among the three countries, it exploits the Web efficiently as shown below:

1. Adapted mean value of international peer inlinks is the largest amongst the three countries.
2. Median national peer inlinks and interconnection rate are larger than those of the UK physics departments.
3. Proportion of national inlinks is the smallest amongst the three countries.
4. In physics, from the UK to Australia, the link propensity is the largest.

- Canadian physics departments

Canadian physics departments' world relative citation impact is smaller than that in the UK.

1. Median external inlinks is the largest amongst the three countries.
2. Median number of national peer inlinks is the largest amongst the three countries.
3. Interconnection rate is the largest amongst the three countries.
4. Proportion of national inlinks is smaller than that of the UK physics departments.
5. Adapted mean value of international peer inlinks is larger than that of the UK physics departments.

- Canadian biology departments

The set of departments has the smallest world citation relative impact amongst the three countries.

1. Median web pages and external inlinks are the largest amongst the three countries.
2. Proportion of national inlinks is the smallest among the three countries.
3. Median number of web pages is larger than that of the physics departments in Canada.
4. From Australia to Canada, biology departments' link propensity is larger than that of the chemistry departments.

- UK chemistry departments

UK chemistry departments' world relative citation impact is smaller than that of the Canadian chemistry departments.

1. Median national peer inlinks and interconnection rate are the largest amongst the three countries.
2. Proportion of national inlinks is the smallest amongst the three countries.

3. Adapted mean number of international peer inlinks is the largest amongst the three countries, and also is larger than that of the UK physics departments.
4. From Canada to UK, chemistry departments' link propensity is the largest.

7.3.3.2 Bad Web Exploitation

- Canadian chemistry departments

The set of departments have the largest world relative citation impact among the three countries.

1. Median web pages and external outlinks are smaller than that of the UK chemistry departments.
2. Median number of national peer inlinks is smaller than the UK chemistry departments.
3. Adapted mean value of international peer inlinks is the smallest among the three countries.
4. Interconnection rate is smaller than that of the UK chemistry departments.
5. Proportion of national inlinks is larger than that of the UK chemistry departments.
6. Median external inlinks is the smallest amongst the three disciplines in Canada.
7. In chemistry, from Australia to Canada, the link propensity is the smallest.

- UK physics departments

UK physics departments have the largest world relative citation impact amongst the three countries. However, its performance on the Web is not as good as expected, although the set of departments have the largest median web pages and create the largest median all and external outlinks among the three countries.

1. Proportion of national inlinks is the largest amongst the three countries.
2. Adapted mean number of international peer inlinks is the smallest amongst the three countries, and also is smaller than that of the UK chemistry departments.
3. In physics, the link propensity from Canada to the UK is the smallest.

8. Discussion

In this chapter, firstly methodological issues are discussed. The apparent outliers from the linear diagrams in chapter 6 are then identified, and reasons for them are tracked down. Finally, the link patterns and anomalies identified in chapter 7 are discussed.

8.1 Methodological Issues

Departmental link analysis involves more problems than university link analysis. Departmental level data collection is more difficult with regard to both web and non-web data. Furthermore, as departments are at a lower aggregation level than whole universities, statistically significant tests are less likely. Some of the methodological issues are discussed in this section.

8.1.1 Critical Issues in Web Data Collection

In addition to the generic data collection problems discussed in chapter 2, there are additional departmental level issues. As discussed in section 3.1, with regard to the preliminary study of the 79 computer science departments in the UK, the identification of departmental domain names is more complicated than that for whole universities. For example, major problems are caused by the identification of old department domain names. Recall that the rationale behind the inclusion of old domain names is to give the departments fairer inlink counts, especially for departments that have just changed their domain names. However, this may also introduce dead target pages (web pages being linked but already disappeared).

The domain names for the physics, chemistry and biology departments in Australia, Canada and the UK are listed in Appendices B1, B4 and B7, C1, C4 and C7 and D1, D4 and D7. To identify the domain names, the departments involved must first be identified. Even this is not a straightforward task. The situation is that the physics, chemistry and biology departments may consist of more than one department (sometimes also described as a school) or research group, institute, laboratory and unit which are classified as physics, chemistry and biology. For example, physics departments may include space and astronomy, condensed matter physics, subatomic physics and theoretical physics departments or research groups. Chemistry may include departments or groups such as inorganic and organic chemistry, and biochemistry. Biology may include departments or groups such as biology, zoology, plants, physiology, genetics and anatomy, and microbiology. The departments in Australia and the UK are classified in the REPP (details see Appendix I) and RAE 2001 schemes separately. For Canada, the departments or research units which received research grants in 2003 were categorized into physics, chemistry and biology according to the Dewey Decimal Classification (DDC, 2003) and the Library of Congress Bibliographies (LCB, 2003). Nevertheless, it is acknowledged that the classification is not exact, because of the existence of multi-disciplinary departments.

The domain or directory names for physics and chemistry departments are fewer in number than those for biology departments. This is because the biology subject has more departments or units classified as biology. For example, the biology department in the University of Manitoba in Canada contains twelve domain names altogether. The physics, chemistry and biology departments in this study in fact are all aggregations of sets of related departments or research units. To make the situation worse, a department may share a homepage with other disciplines. For example, http://www.scieng.flinders.edu.au/cpes/ is the homepage of the school of chemistry, physics and earth science and science engineering from Flinders University in Australia. There is not a separate homepage for the chemistry or physics departments; in this case both the physics and chemistry departments have the same directory name. This may not be a problem when counting links from their peers, as departments are more likely to be targeted by the same type (Harries et al., 2004). However, when counting the number of links from large web areas such as the whole Web, the results certainly will be inflated, as the inlinks to other disciplines are also included.

The number of inter-departmental links made amongst a set of departments in the same country is small compared with the external outlinks made from the departments to the universities that host the same set, as shown in table 6.69. In other words, departments link more outside their discipline than within their discipline, at least within their own country. The sparse number of inter-departmental links is certainly not a good thing for conducting a statistical test. However, everything has two sides. The relatively small number of links between departments makes it possible to visit all the target pages, and identify the nature of hyperlinks among peer departments.

8.1.2 Critical Issues for non-Web Data Collection

The non-web data collection for departments also has proved to be more troublesome than that for universities. Numbers of academic staff members for whole universities can easily be found in the International Handbook of Universities (IAU, 2003a) or World Higher Education Database (IAU, 2003b). The UK departments' numbers of academic staff members are included in the RAE 2001 submission. However, no official data for the departments' academic staff members in Australia and Canada has been found. The number of academic staff members for the departments had to be manually identified from the departments' web sites. As each department has a different web design, it is not feasible to collect this data automatically on the Web.

With regard to research measures, citation counts are collected from ISI's Web of Science in a semi-automatic way. This is because that the citation counts are the most relevant data set to compare with link counts in the correlation tests. On the other hand, publicly accessible research performance reports are normally organized at university level. Thus citation counts are always a good candidate to represent departments' research performance when other measures not available. For example, Smith & Thelwall (2002) used the Research Quantum (DETYA, 2001) to represent Australian universities' research performance. However, this data is not available at departmental level.

In addition to citation counts, RAE 2001 provides the research ratings for the UK departments. However, the NSERC research grants for Canadian departments are organized both by field and university, but not by department (NSERC, 2003). In this study, the data has to be reorganized by departments.

8.1.3 Critical Issues in the Identification of Motivations for Departmental Interlinks

8.1.3.1 A Comparison with Other Approaches

The motivations for citations between journal papers have been investigated for more than three decades without a clear answer (Case & Higgins, 2000). This is because the process of creating a citation is both private and complex, even directly interviewing with authors does not necessarily give correct answers, as they may forget the original reasons for their citations. It is more difficult to identify motivations for creating hyperlinks. Firstly, it is not always possible to identify the creator of a web page. Even if a webmaster is identified, he or she may not be the person who makes decisions about whether to add or remove hyperlinks from web pages. Secondly, the content of web pages is not peer reviewed, thus there is no quality guarantee. Finally, web pages are dynamic by nature and can disappear at any time. Obviously a direct survey of creators of web pages is not a plausible method for any large scale analysis. The investigation has to be concentrated on link contexts. Recall that some researchers have focused on link targets (Thelwall, 2002g; Thelwall et al., 2003c), some on the link contexts of source and target pages (Wilkinson et al., 2003), some on link source and target pages (Harries et al., 2004) and others on all the different components: link sources, targets and the links themselves (Bar-Ilan, 2004b; Bar-Ilan, 2004a).

As the numbers of links among departments are highly skewed and small compared with those at the university level, it is not as easy to get a representative sample as for university links (Wilkinson et al., 2003). In this study, all target pages are visited to ascertain their types, and whether they are still there. One critical issue is that only one person carried out the categorization process in this project, while other studies normally have more than one coder and the inter-coder consistency can be used to prove the reliability of the results. However the definitions for each category are relatively straightforward, and so this is not believed to be a serious problem.

Although the classification scheme applied in this project is simpler than those conducted by Wilkinson et al., Harries et al. and Bar-Ilan, by classifying the whole population of the target pages, it accurately illustrates the disappearance rate of departmental target pages; and by tracing back to some special target pages' source pages, the nature of inter-departmental hyperlinks is illustrated more in context than would be possible through merely classifying the target pages.

8.1.3.2 Difficulties in Categorizing the Target Pages

The categorization is not always clear-cut. For example, PhD students are not academic staff members, although they are different from undergraduate students. Some universities have categorised them as research staff together with research fellows, assistants and associates. In this study, homepages of PhD students are grouped into 'homepages of staff members' and this is the same as Harries et al.'s approach.

The 'resource' category includes a variety of different contents, such as course related information, technical documentation, databases, link lists of staff members or research units, student societies, national and international organizations and conferences, online journals, research papers or job information. Student societies, national or international organizations and conferences are mostly linked through their homepages, however, they neither represent the department nor a research group within the department where they are hosted, and they contain some useful information and are therefore put in the 'resource' category.

The most obvious 'disappeared' pages are those that no longer exist. This might be caused by a staff member moving to another university; out of date information no longer supported by a server, or the whole department's url structure being reorganized. Web pages might also be redirected to new positions. The 'disappeared' category also involves some problems. For example, a staff homepage http://www.ncl.ac.uk/chemistry/research/btg.html from the chemistry department in University of Newcastle is now redirected to the nature science department, which is combined by the chemistry and physics departments (http://www.ncl.ac.uk/nsci/). The original chemistry department's homepage (http://www.ncl.ac.uk/chemistry/) no longer exists. In this case the staff homepage mentioned above is classified as 'disappeared'. Another staff homepage from the Hull University, http://www.hull.ac.uk/chemistry/staff/jdc.html, however, is redirected to http://www.hull.ac.uk/chemistry/academic_staff.php?id=jdc. This page is still regarded as a staff homepage because the redirected pages still point to that staff member's homepage. Homepages of departments or groups are less likely to disappear. Firstly, accordingly to Koehler (2002), higher level urls are more stable than lower ones. Even if their urls change, they are more likely to be redirected to the new positions. The urls for resources or homepages of staff members which are deeper in the directory structure, are more dynamic and once they disappear, they are less likely to be redirected to the exact new positions as more effort is needed.

No recreational pages have been found in this study. In a previous study by Wilkinson et al. (2003), the links for recreational purposes between universities in the UK were 6.5%. The fact that no recreation target pages have been found in this study may suggest that at the departmental level the interlinking is more consistently serious than that at the university level. In an earlier study by Thelwall (2001b), apart from the 'disappeared' and 'others' categories in this study, the rest were all regarded as research related. In that study, even the homepages of departments were regarded as research related web pages. Thelwall argued that they contain more academic related resources than homepages of whole universities.

It is hard to generalize the exact motivations for departmental interlinks, as there is no uniformity in creating web pages for departments. For example, homepages of staff members may contain research related contents, or just telephone numbers and e-mail addresses. Nevertheless, the categorization for each set of departments' target pages has shed some light on the reasons for departmental interlinking.

8.2 Outlier Investigations

Strong linear trends are shown in the diagrams of chapter 6, especially those showing file or external inlinks and research productivities. In order to better understand the connection between research and link data, departments that do not fit the trends will be identified, and reasons why the departments receive more or less inlinks than expected will be tracked down.

8.2.1 Apparent Outliers for Australian Departments

8.2.1.1 Outliers in figure 6.1

- Australian Defence Force Academy, University of New South Wales (714, 1)
- Monash University (265, 17)
- Flinders University (428, 28)
- Macquarie University (864, 65)

As listed in table 8.1, the physics department in the Australian Defence Force Academy from the University of New South Wales has a smaller web site than the median value. Either the department makes less use of the Web or it is unable to be crawled properly by SocSciBot3. It attracts less inlinks relatively to the citations it received. The physics departments in Monash and Macquarie universities have larger than median web sites and attract more median inlinks from their peers in Australia. In this context, Flinders University is an exception. It has less than median web pages, but attracts more than median inlinks.

The physics department from the University of Queensland linked to http://www.physics.mq.edu.au/~acols/ 18 times. This web page, hosted by Macquarie University, is about the Australian Conference on Optics, Lasers and Spectroscopy 2001 and had disappeared when visited.

Table 8.1. Web pages for the outliers in figure 6.1 and relevant median values

University name	Web page number	Median type	Median value
Australian Defence Force Academy, University of New South Wales	43	Median web pages	327.5
Monash University	1015	Median citations	450.5
Flinders University	281	Median peer inlinks	14
Macquarie University	5063		

8.2.1.2 Outliers in Figure 6.2

- Australian Defence Force Academy, University of New South Wales (714, 45)
- Monash University (265, 408)
- University of Melbourne (2670, 4170)

In table 8.2, the physics department from the Australian Defence Force Academy, University of New South Wales has a smaller than median web site and attracts fewer than median inlinks, although the citations received by the department is larger than median value. Either it creates a smaller number of web pages or it is not indexed well by AltaVista. The physics department of Monash University has more than the median number of web pages and attracts more than median inlinks, even if its citations received is smaller than median value. The physics department in Monash University shares the homepage http://www.spme.monash.edu.au with other disciplines. This is another reason for the department to have a larger web site and attract more inlinks from the whole Web. The physics department in the University of Melbourne has a larger than median web site, and attracts more than median inlinks from the Web. The department's number of citations received is also larger than the median value.

Table 8.2. Web pages for the outliers in figure 6.2 and relevant median values

University name	Web page number	Median type	Median value
Australian Defence Force Academy, University of New South Wales	8	Median web pages	143.5
Monash University	395	Median citations	450.5
University of Melbourne	11281	Median external inlinks	101.5

8.2.1.3 Outliers in Figure 6.5

- University of Western Australia (2028, 1)
- University of Adelaide (1677, 1)
- Queensland University of Technology (995, 0)
- Macquarie University (370, 9)
- Australian Defence Force Academy, University of New South Wales (235, 5)

As illustrated in table 8.3, the chemistry departments from the universities of Western Australia, Adelaide and Queensland University of Technology all have smaller than median web pages. Even if their citation counts are all larger than the median value, they all attract fewer than median inlinks from their peers. Macquarie University has a larger than median web site, and even if its citation counts is smaller than the median value, it attracts more than median inlinks. However, the chemistry department in Australian Defence Force Academy, University of New South Wales has a smaller than median web site, but it attracts more than median inlinks. This departments' number of citations received is also smaller than the median value. 4 out of the 5 inlinks are from the Australian National University, where two are from a visiting research fellow who is also a senior research associate from the chemistry department in the Australian Defence Force Academy and two others are from two chemistry departments' link lists.

Table 8.3. Web pages for outliers in figure 6.5 and relevant median values

University name	Web page number	Median type	Median value
University of Western Australia	424	Median web pages	220
University of Adelaide	172	Median citations	442
Queensland University of Technology	227	Median peer inlinks	2
Macquarie University	496		
Australian Defence Force Academy, University of New South Wales	69		

8.2.1.4 Outliers in figure 6.6

- University of Tasmania (1217, 2646)
- Queensland University of Technology (995, 31)
- Australian Defence Force Academy, University of New South Wales (235, 358)

The number of web pages for Queensland University of Technology is only 1, which is obviously wrong. Even if it received larger than median citations, it still attracts fewer than median external inlinks. The number of web pages for the chemistry department in Australian Defence Force Academy, University of New South Wales is much larger than the median value, and it also receives more than median external inlinks from the Web. The chemistry department in the University of Tasmania shares its homepage with departments from other disciplines (http://www.scieng.utas.edu.au/scieng/). The large number of inlinks from the Web for this department might be attracted by other disciplines, as it is not an outlier in figure 6.5, which shows the number of inlinks from its peers in Australia. As shown in table 8.4, the number of web pages for this department is only 3, which is obviously wrong. Nevertheless, I could not find out the reason for this, as AltaVista does not supply the same advanced facility now.

Table 8.4 Web pages for the outliers in figure 6.6 and relevant median values

University name	Web page number	Median type	Median value
University of Tasmania	3	Median web pages	116
Queensland University of Technology	1	Median citations	442
Australian Defence Force Academy, University of New South Wales	3181	Median external inlinks	100

8.2.1.5 Outliers in figure 6.9

- University of Adelaide (4437, 0)
- University of Canberra (340, 7)
- University of Wollongong (786, 5)

As illustrated in table 8.5, the biology department in the University of Adelaide has less than median web pages and receives 0 inlinks, although it received 4437 citations which is much larger than the median value. Biology departments in the University of Wollongong and the University of Canberra have larger than median web sites and attract more inlinks, even if the numbers of citations received by them are all smaller than the median value.

Table 8.5. Web pages for the outliers in figure 6.9 and relevant median values

University name	Web page number	Median type	Median value
University of Adelaide	193	Median web pages	267
University of Canberra	10391	Median citations	1067.5
University of Wollongong	417	Median peer inlinks	1

8.2.1.6 Outliers in figure 6.10

- University of Canberra (340, 932)
- Queensland University of Technology (435, 625)
- University of Adelaide (4437, 36)

The biology departments in the University of Canberra and Queensland University have larger than median web sites, as shown in table 8.6 and attract more inlinks from the Web, even if they receive smaller number of citations than the median value. The biology department of the University of Adelaide has smaller web site than the median value, and attracts fewer than median inlinks, even if its number of citations received is larger than the median value.

Table 8.6. Web pages for the outliers in figure 6.10 and relevant median values

University name	Web page number	Median type	Median value
University of Canberra	391	Median web pages	221
Queensland University of Technology	1315	Median citations	1067.5
University of Adelaide	230	Median external inlinks	80.5

8.2.2 Outliers for Canadian Departments

8.2.2.1 Outliers in figure 6.13

- Queen's University (299870, 156)
- University of Quebec (741381, 3)
- University of Montreal (883307, 5)

The physics department in Queen's University has a larger than median web site, as listed in table 8.7 and attracts more than median inlinks. The department received altogether 156 inlinks from its peers, and 88 are received by the Sudbury Neutrino Observatory (http://www.sno.phy.queensu.ca/). Thus, one unit within a department attracts the majority of the inlinks. Those in the universities of Quebec and Montreal have smaller than median web sites and attract fewer than median inlinks, even if both departments received more than median research grants. Another reason might well be the French language used on their web sites. The same was found for whole universities' link study in Canada (Vaughan & Thelwall, 2005).

Table 8.7 Web pages for the outliers in figure 6.13 and relevant median values

University name	Web page number	Median type	Median value
Queen's University	2456	Median web pages	568
University of Quebec	374	Median grants	173000
University of Montreal	295	Median peer inlinks	16

8.2.2.2 Outliers in figure 6.14

- University of Montreal (883307, 117)
- University of British Columbia (675308, 4830)
- Memorial University of Newfoundland (58800, 1134)

The physics department in the University of Montreal was again an outlier in figure 6.14. The underlying reason might be the same as found in figure 6.13, the smaller than median number of web pages indexed by AltaVista and the site being in French. The physics departments in the University of British Columbia and Memorial University of Newfoundland have larger than median web sites, as illustrated in table 8.8 and attract more than median inlinks from the Web.

Table 8.8. The number of web pages for the outliers found in figure 6.14

University name	Web page number	Median type	Median value
University of Montreal	42	Median web pages	603
University of British Columbia	10795	Median grants	173000
Memorial University of Newfoundland	2599	Median external inlinks	467

8.2.2.3 Outliers in figure 6.17

- University of Montreal (620331, 8)
- McGill University (985619, 24)
- University of Guelph (465613, 428)
- Dalhousie University (206571, 51)

The chemistry departments in Universities of Montreal and McGill are both from French speaking Quebec. This might be the reason why they received fewer inlinks from their peers, as the research grants they received are both larger than the median value. Although McGill University's web site is mainly in English, it still receives fewer than median inlinks, as shown in table 8.9. The chemistry department in the University of Guelph has larger than median web site, however, the immediate reason for it receiving larger than median number of inlinks is that the chemistry department in the University of Calgary http://www.cobalt.chem.ucalgary.ca/ziegler/educmat/chm386/tutor386.htm mirrored an online tutorial written by Dan Thomas from the University of Guelph. At the bottom of each page in that tutorial, there is a link back to the author's homepage (http://www.chembio.uoguelph.ca/thomas/thomas.htm). There are altogether 392 such links. This is an anomaly, as the links are all from the same motivation.

Table 8.9 Web pages for the outliers in figure 6.17 and relevant median values

University name	Web page number	Median type	Median value
University of Montreal	346	Median web pages	279.5
Mcgill University	351	Median grants	209375.5
University of Guelph	2754	Median peer inlinks	8.5
Dalhousie University	259		

8.2.2.4 Outliers in figure 6.18

- University of Alberta (711290, 2110)
- University of Guelph (465613, 1827)
- McMaster University (363218, 1357)
- University of Montreal (620331, 71)
- McGill University (985619, 135)

Again in figure 6.18 the chemistry departments in the Universities of Montreal and McGill appear to be outliers. They both received fewer than median inlinks, although they both receive more than median research grants, as shown in table 8.10. The web site of University of Montreal has larger than median web pages, the reason might be being in French. The number of web pages for Universtiy of McGill is only 10, obviously the department is not indexed well by AltaVista. The chemistry departments in the University of Alberta, Guelph and McMaster all have larger than median web sites (see Table 8.10) and receive more than median inlinks from the Web. Their research grants are all larger than the median value.

Table 8.10 Web pages for the outliers in figure 6.18 and relevant median values

University name	Web page number	Median type	Median value
University of Alberta	2435	Median web pages	170.5
University of Guelph	3390	Median grants	209375.5
McMaster University	1786	Median external inlinks	87.5
University of Montreal	216		
McGill University	10		

8.2.2.5 Outliers in figure 6.21

- Université du Québec à Montréal (493908, 0)
- University Laval (620853, 1)
- McMaster University (113850, 12)

Université du Québec à Montréal and University Laval are both from Quebec. Although the biology department in the University of Laval has a larger than median web site, as shown in table 8.11, and received larger than median research grants, it receives fewer than median inlinks from its Canadian peers. The reason may be being in French. Although the biology department in Université du Québec à Montréal received more than median research grants, its number of web pages is smaller than the median value, and it receive fewer than median peer inlinks. McMaster University has a larger than median web site and receives more than median peer inlinks, even if its research grants received is smaller than the median value.

Table 8.11 Web pages for the outliers in figure 6.21 and relevant median values

University name	Web page number	Median type	Median value
Université du Québec à Montréal	174	Median web pages	584.5
University Laval	2272	Median grants	269703.5
McMaster University	899	Median peer inlinks	3

8.2.2.6 Outliers in figure 6.22

- University of Guelph (1579467, 91)
- McMaster University (113850, 1003)
- Université du Québec à Montréal (493908, 0)
- University Laval (620853, 176)

Université du Québec à Montréal and University Laval again are outliers in figure 6.22 as in figure 6.21 for having smaller than median web sites and being in French, even if their research grants received are smaller than the median value. The biology department from McMaster has a slightly larger than median web site, as shown in table 8.12. Although its research grants are smaller than the median value, it attracts more than median external inlinks.

Table 8.12 Web pages for the outliers in figure 6.22 and relevant median values

University name	Web page number	Median type	Median value
University of Guelph	62	Median web pages	308.5
McMaster University	334	Median grants	269703.5
Université du Québec à Montréal	6	Median external inlinks	102.5
University Laval	265		

8.2.3 Outliers for UK Departments

8.2.3.1 Outliers in figure 6.25

- Imperial College London (698.8, 0)
- University of Manchester (354, 0)
- Queen's University Belfast (330, 0)
- University of Glasgow (244, 108)
- Open University (60, 38)

Both the link structures of the physics departments in Imperial College London and Queen's University Belfast could not be extracted from their parent universities' link structure files. They were all crawled separately again. This might be the reason why they get fewer inlinks from their peers in the UK, as they are invisible even from their own universities, even if their research productivities are both larger than the median value. The numbers of web pages for physics departments in the Imperial College London, the University of Manchester and Queen's University Belfast are smaller than the median value as shown in table 8.13. They receive fewer peer inlinks than the median value. Although the Open University has smaller than median web pages, its research productivity is also smaller than the median value; the number of inlinks is larger than the median value. The physics department in University of Glasgow has a larger than median web site, and attracts more than median peer inlinks.

Table 8.13 Web pages for the outliers in figure 6.25 and relevant median values

University name	Web page number	Median type	Median value
Imperial College London	114	Median web pages	820
University of Manchester	77	Median research productivities	159.35
Queen's University Belfast	280	Median peer inlinks	4.5
University of Glasgow	1703		
Open University	499		

Two staff members' homepages http://physics.open.ac.uk/~ajnorton/, and http://physics.open.ac.uk/~chaswell/ from the Open University's physics department are linked by the institute of astronomy hosted by Cambridge

University 16 times altogether for the same reason, as occurred in a physics scientists' list. This is the immediate reason why the Open University is an outlier in figure 6.25.

http://www.physics.gla.ac.uk/gwg/, the homepage of institute for gravitational research in the University of Glasgow has been linked 19 times by the physics department in Cardiff University. A set of slides hosted by the physics department in University of Glasgow is targeted by the physics department in University of Strathclyde. By the time the target pages were visited, they had already disappeared. However, according to the content of the URLs, one could guess that they were a set of 26 slides, each slide having a different url, the first being
http://www.physics.gla.ac.uk/lis/show/sld001.htm, and the last being
http://www.physics.gla.ac.uk/lis/show/sld026.htm. The 26 links are from the same motivation, to point to the set of slides. If the 45 inlinks were removed, then the physics department of Glasgow University would not be an apparent outlier in figure 6.25.

8.2.3.2 Outliers in figure 6.26

- Imperial College London (698.8, 3)
- Queen's University Belfast (330, 1)
- University of Exeter (120, 4840)

The physics department in Imperial College London and Queen's University Belfast are again found to be outliers in figure 6.26. They have fewer than the median web pages according to the link data collected by AltaVista, as shown in table 8.14, even if their research productivities are both larger than the median value, they receive fewer than medina external inlinks. The number of median web pages for the physics department in the University of Exeter is huge at 5,994, which is much larger than the median value. This may be the reason why this department attracts much larger than median inlinks from the whole Web, even if its research productivity is smaller than the median value.

Table 8.14 Web pages for the outliers in the figure 6.26 and relevant median values

University name	Web page number	Median type	Median value
Imperial College London	22	Median web pages	528.5
Queen's University Belfast	370	Median research productivities	159.35
University of Exeter	5994	Median external inlinks	238

8.2.3.3 Outliers in figure 6.29

- Nottingham Trent University (30, 81)
- University of Southampton (227.7, 10)
- University of Sussex (198, 5)
- University of Sheffield (180, 223)

In table 8.15, the number of web pages for the chemistry department in Nottingham Trent University is the same as the median value, and the number of inlinks is 81 (median value: 82). However, its research productivities is only 30 which is much smaller than the median value (140), so it appears to be an outlier in figure 6.29. The chemistry department in Sheffield University has a larger web site than the median value, and receives more than median inlinks. The chemistry departments in the University of Southampton and Sussex have fewer web pages than the median value, and they both attract fewer than median inlinks, even if their research productivities are both larger than the median value.

Chemistry Staff members' homepages in the University Sheffield are linked by the chemistry department in Cambridge University 92 times. This is caused by the web page http://www.ch.cam.ac.uk/c2k/people/StaffList/StaffList00.html hosting a UK chemists list. The chemistry department in Oxford made links to the site http://www.shef.ac.uk/chemistry/web-elements/ 86 times. If these links were removed, the University of Sheffield would not have been an outlier in figure 6.29.

Table 8.15 Web pages for outliers in figure 6.29 and relevant median values

University name	Web page number	Median type	Median value
Nottingham Trent University	426	Median web pages	426
University of Sheffield	852	Median research productivities	140
University of Sussex	184	Median peer inlinks	82
University of Southampton	700		

8.2.3.4 Outliers in figure 6.30

- Queen Mary University of London (61.5, 5323)
- University of Sheffield (180, 7983)
- University of Durham (255.5, 46)
- University of Sussex (198, 40)
- University of Southampton (227.7, 199)

In table 8.16, although the chemistry departments in the universities of Durham and Sussex both have larger than median research productivities, they attract fewer than median inlinks from the whole Web. The chemistry department in Queen Mary University of London has a larger than median web site and attracts more than median inlinks from the whole Web, even if its research productivity is smaller than the median value. The chemistry department in Southampton has a larger than median web site, and attracts more than median inlinks from the Web. The chemistry department in the University of Sheffield has a larger than median web site, and it receives a huge number of inlinks (7983) which is much larger than the median value. Although the department's research productivity is also a little bigger than the median value, it is difficult to reply why the department attract such a large number of inlinks.

Table 8.16 Web pages for outliers in figure 6.30 and relevant median values

University name	Web page number	Median type	Median value
Queen Mary University of London	4314	Median web pages	329
University of Sheffield	549	Median research productivities	140
University of Durham	309	Median external inlinks	176
University of Sussex	329		
University of Southampton	712		

8.3.2.5 Outliers in figure 6.33

- University College London (518.4, 2129)
- Birkbeck University of London (174.6, 105)
- University of Edinburgh (918.9, 8)

As listed in table 8.17, the biology departments from University College London and Birkbeck College both have larger web sites than median value. The research productivity of biology department in Birkbeck is nearly the same as the median value, it receive 44 inlinks from University of College London, 14 from Cambridge University. The immediate cause for so many inlinks to University College London, however, is the biology department hosting a lot of databases heavily linked by other UK biology departments, especially by the biology department in Cambridge University which created 1,720 links to the databases. Details of the databases are listed in Appendix H7. Although both the number of web pages and research productivity for the biology department in the University of Edinburgh are larger than the median values, it attracts fewer peer inlinks than the median value. It seems that the department needs to pay some attention to enhance its web visibility.

Table 8.17 Web pages for outliers in figure 6.33 and relevant median values

University name	Web page number	Median type	Median value
University College London	2445	Median web pages	576
Birkbeck, University of London	16946	Median research productivities	172.5
University of Edinburgh	1455	Median peer inlinks	3

Once the two apparent outliers, the biology departments from the University College London and Birkbeck College are removed, the correlation coefficient values are more significant, as listed separately in tables 7.18 and 7.19, than those in tables 6.57 and 6.58. It shows how individual department's link behaviour can have a large impact on a set of departments in a country.

Table 8.18. Pearson correlations between link (from SocSciBot) and research (RAE) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=51)

Link measure	File	Directory	Domain	Department
Inlinks	0.777**	0.759**	0.788**	0.756**
Outlinks	0.495**	0.492**	0.643**	0.668**
WIFs (staff)	0.542**	0.529**	0.520**	0.485**
WUFs (staff)	0.222	0.209	0.437**	0.416**

Table 8.19 Spearman correlations between link (from SocSciBot) and research (RAE) measures. (* = significant at the 5 % level, ** = significant at the 1% level, n=51)

	File	Directory	Domain	Department
Inlinks	0.796**	0.795**	0.797**	0.785**
Outlinks	0.660**	0.654**	0.703**	0.695**
WIFs (staff)	0.631**	0.626**	0.585**	0.518**
WUFs (staff)	0.488**	0.483**	0.494**	0.485**

8.2.3.6 Outliers in figure 6.34

- University College London (518.4, 14470)
- Birkbeck College, University of London (174.6, 5666)
- University of Edinburgh (918.9,693)

The three outliers found in figure 6.34 are exactly the same as those found in figure 6.33. The reason might be the same. As shown in table 8.20, the biology departments from the University College London and Birkbeck College are again found to have more than median web pages, and attract more than median number of inlinks from the whole Web. The biology department in Edinburgh has larger than median web pages, and receives larger than median inlinks, however, compared with the large research productivity the department has, the number of inlinks is relatively small.

Table 8.20. Web pages for the outliers in figure 6.34 and relevant median values

University name	Web page number	Median type	Median value
University College London	110744	Median web pages	543
Birkbeck University of London	7755	Median research productivities	172.5
University of Edinburgh	1107	Median external inlinks	219

8.2.4 Outlier Analysis Summary

Through the analysis above, departments with larger web sites tend to attract more inlinks both from their peers in the same country and elsewhere on the Web, while departments with smaller web sites tend to attract fewer inlinks. In order to enhance their web visibility, departments should consider creating more web publications. Departments that share the same homepages with other disciplines may receive a biased larger number of inlinks from the whole Web, as the number of inlinks from other disciplines has been included. The departments in Quebec are isolated from both their peers in Canada and the whole Web.

Creating web pages not only in French but also in English might be a useful way for them to change this situation.

8.3 Linking Differences Along National and Disciplinary Lines

As discussed in section 6.3.2, the departmental interlinking may reflect informal scholarly communication amongst departments. Different link patterns identified along country and disciplinary lines may well be influenced by formal scholarly communication. The underlying reasons for the patterns will be discussed in this section. Departments that do not fit well into the patterns will also be discussed.

8.3.1 National Comparisons

The results show that for general web use, UK departments are the best, followed by Canada and then Australia. This finding is in parallel with the countries' research performances. Table 8.21 lists relative citation impact compared to world average in each field in percentage terms ('world relative citation impact' for simplicity), for each set of departments in the three countries from ISI (2003). England's result is used to represent the UK, and biology and biochemistry is chosen to represent biology as no biology category is present. The UK's world relative citation impacts for the three disciplines are nearly all the largest, those for Canada are the second and Australia the smallest. With one exception: the value for UK chemistry is smaller than that of Canada.

Canadian chemistry departments and UK physics departments are relatively invisible on the Web with regard to their research performances. Despite the fact that both departments have the largest world relative citation impact amongst the three countries, Canadian chemistry departments receive the smallest number of inlinks from different web areas, and the UK physics departments receive fewer inlinks from different web areas than those of Canada. The two departments may need to enhance their web profiles in order to match their research performances.

With regard to national peer interlinks, the Canadian departments interconnect the best with their national peers, and the least with their international peers. The Canadian departments might be biased in this study as to their ability to attract international peer inlinks: according to the collaboration patterns identified through co authorship analysis in bibliometrics (Glänzel, 2001; Glänzel & Schubert, 2001), the UK and Australia have closer collaboration than with Canada. In this project, the UK and Australian departments prefer to outlink to each other's country domain than the Canadian (ca) domain. Should other countries, such as the United States be included in this study, a different conclusion might have been drawn for Canadian departments.

Although the UK physics departments have the largest median web pages, creating the largest median outlinks, receiving the largest median inlinks from different web areas and with the best research performance, the set of departments' median national peer inlinks and adapted mean international peer inlinks are both the smallest amongst the three countries. They need to pay particular attention to enhance their visibilities on the Web to their peers. On the other hand, the Australian physics departments' visibility from their national and international peers is the best amongst the three countries, despite their smallest world relative citation impacts.

Table 8.21 The world relative citation impact (percentages) for Australia, Canada and the UK in Biology, Chemistry and Physics, from ISI (2003)

Discipline	Australia	Canada	UK
Biology	-6	3	15
Chemistry	8	27	24
Physics	9	18	29

8.3.2 Disciplinary Comparisons

For general web use, physics departments are the best, followed by biology and then chemistry. For peer interlinking, a pattern identified for both national and international peers is that physics departments are the best, followed by chemistry and then biology. Physics departments perform the best on the Web amongst the three disciplines in each country, the chemistry departments the least on general web use, while biology departments the least on peer interlinking. The subtle difference between chemistry and biology departments is very interesting, as there is more similarity between chemistry and biology than with physics (e.g. biochemistry).

This favourable showing of physics departments is not surprising. In the study of (Harries et al., 2004), physics web pages make more intra-subject links than those in maths and sociology. As discussed in section 6.3.2, physics departments prefer to link the most to useful resources suggesting that physics is particularly well supplied with online information. In table 6.70, the physics departments' correlations between external inlinks and research productivities are the largest in each country. The physics departments tend to make the most use of the Web, just as physicists tend to distribute their research quickly and widely in formal scholarly communications (Kling & McKim, 2000). For example, it is a common convention for them to put their preprints online. The situation is different for scientists in chemistry and biology. The different linking behaviours found in this project for different disciplines suggest that link patterns may well mirror offline phenomena. Garfield (1999) stated that 'All citation studies should be normalized to take into account variables such as field, or discipline, and citation practices.', and this can now be seen to apply to the Web, even for similar disciplines in hard sciences. This also gives empirical evidence to support Kling & McKim (2000) who have argued that 'the World Wide Web can be adopted and used by different fields in dramatically different ways.'

The set of UK chemistry departments does not fit the general patterns. It receives the largest median national peer inlinks and adapted mean international peer inlinks, and its interconnection rate is the largest both along country and disciplinary lines. Its proportion of national inlinks is the smallest in chemistry amongst the three countries. From Canada to the UK, the link propensity of chemistry departments is the largest. This set of departments has very efficient web profiles with regard to both general web use, and peer interlinking (both national and international). In section 6.2.4, a UK chemists list identified in Cambridge University causes a huge number of links pointing to other UK chemistry departments. This is one of the reasons why the set of departments performs so well on national peer interlinking.

Canadian biology departments' performance on the Web is better than expected, as discussed in section 7.3.3, it does not fit the general pattern and is listed under 'good web exploitation'. The median web pages from SocSciBot3 for the set of departments is larger than that of the physics departments. The proportion of national inlinks for the set of departments is smaller than that of the physics departments. The median web pages and external outlinks are larger than those for the UK biology departments from SocSciBot3. From Australia to Canada, the link propensity to Biology departments is larger than that of the chemistry departments.

8.3.3 Summary

Apart from some exceptions, link patterns have been identified both along national and disciplinary lines.

8.3.3.1 International Patterns

There are three types of international pattern: general web use, national peer interlinking and international peer interlinking. With regard to general web use, the link patterns have been found to reflect formal scholarly communication. UK departments perform the best in general web use and have the largest world relative citation impact amongst the three countries, followed by Canadian and then Australian departments.

The differences between national and international peer interlink patterns may be caused by the small range of countries included in this project. The UK and Australian departments have been found to outlink to each other's country domain more than to Canadian (ca) domain. Canadian departments interconnect the best with their national peers, however, the worst with their international peers. Canadian departments may be biased in their ability to attract international peer inlinks.

8.3.3.2 Disciplinary Patterns

There are two types of disciplinary pattern: general web use, and peer interlinking (both national and international). Physics departments are found to perform the best in both cases. Chemistry departments make the least general web use, while biology departments perform the least on peer interlinking.

The difference in link patterns along disciplinary lines is also a reflection of informal scholarly communication. Physicists are more willing to distribute their research outputs, while chemists and biologists are more conservative. The different link patterns between chemistry and biology departments with respect to general web use and peer interlinking may reflect that the two disciplines are more similar than with physics.

In summary, link patterns are different for the three similar disciplines in the three similar countries. However, the differences are not independent, they are reflections of off-line phenomena.

9. Conclusions and Future Work

Despite the difficulties encountered with both web and non-web data collection, and the sparseness of interlinking amongst departments, the departmental interlinking investigation in this project has been proved not only to be plausible, but also productive. Not only are significant correlations found between link and research measures, but also different link patterns have been identified along country and disciplinary lines. It is known that very different disciplines have different web profiles. In this project, the three similar departments were found to employ the Web differently with regard to both general web use and peer interlinking. Almost all previous studies of departmental interlinking have had a national focus, as an international one is more difficult to achieve. In this project, the same types of departments in the three countries are found to use the Web differently, and the differences are consistent for the three disciplines; three disciplines use the Web differently in each country, and the differences are consistent for the three countries. Disciplinary and international differences are important for web link researchers, in order to take variations into account when designing procedures and methods in future research, as is currently the case with the more mature similar field of citation analysis.

9.1 Validity of the Departmental Link Data

The significant correlations together with the categorization of target pages serve as the answer to the first hypothesis H1 in section 4.1.2 'Links to departments associate with research in terms of a) significant correlations with existing research measures and b) a majority of links having some association with research.' The significant statistical associations between links and research measures are very useful evidence for the validity of the departmental link data. However, significant correlations do not prove that one is the cause of the other, but the categorization of target pages for the departments is used to further identify what departmental interlinking signifies. Apart from around 20% 'disappeared' target pages for these departments, the majority of the rest are homepages owned by departments, research groups and academic staff, or resources that are useful to the source departments. This suggests that departmental interlinking is predominantly academic-related.

9.1.1 Link-Research Correlations

Significant correlations have been discovered between various link counts and different research measures in this study. Through this, the associations between departmental link counts and research performances are tested from various aspects. Generally, the correlation coefficient values between link counts and research productivities are larger than those between WIFs and research averages. This is because of the size effect, where larger web sites attract more inlinks.

9.1.1.1 Inlinks to a set of departments from the whole Web correlate with departmental research measures significantly

The correlations between external inlinks and research productivities are nearly all significantly at the 1% level for each set of departments. However, the correlations between external WIFs (staff) and research averages are less significant.

9.1.1.2 National Peer Inlinks or Outlinks Correlate with Departmental Research Measures Significantly

The correlations between peer inlinks or outlinks with different ADMs and research productivities are nearly all significantly at the 1% level. After both the links and research productivities are divided by academic staff members, the WIFs (staff) still correlate with research averages, although the correlation coefficient values are less significant. This shows that although size of department is a reason to attract peer inlinks, after the size effects are removed, research performance is still a reason for departments to attract peer inlinks.

9.1.1.3 Inlinks to a Set of Departments from Some Important Domains Correlate Significantly with Departmental Research Measures

The correlations between inlinks from different web domains and research productivities are nearly all significant at the 1% level for each set of departments in the three countries. Those between relevant WIFs (staff) and research averages are less significant.

Hardly any significant correlations have been found between WIFs (with web pages, directories, domains and departments as denominators) and research averages, not even for Australian physics departments, which have the highest correlation between links and research productivities, and between WIFs (staff) and research averages in this project. Highly rated academic staff members tend to create more web pages and receive more inlinks from peer departments or different web areas. However, their individual web pages do not necessarily have a higher impact.

Regarding the investigation of outliers in linear diagrams for the departments in Section 8.2, departments that receive extraordinarily many inlinks are found to have larger web sites, while those with smaller web sites receive fewer inlinks. Departments in Quebec (Canada) are found to be isolated both from their peers and different domains. Departments that want to enhance their web visibility should consider creating web publications as much as possible. Departments in Quebec may consider creating web pages not only in French but also in English in order to change the current situation.

9.1.2 Categorization of Target Pages

The majority of the departmental target pages are categorized as homepages of departments, research groups or staff members, or resources, and only around 20% of the target pages are categorized as 'disappeared'. This suggests that the interlinking among the departments is academic related. The lack of journal papers in the target pages shows that the interlinking among the departments is different from that between citations among journal papers. The fact that no recreational target page was found in this study shows that departmental interlinking is more serious than that of university interlinking, although the numbers of links to departments are smaller than that to universities.

Physics departments in the three countries target the 'resource' type of web pages more than chemistry and biology departments. This shows that physics departments tend to make the most general use of the Web amongst the three disciplines, just as physicists are more willing to distribute their research quickly and widely in formal scholarly communication. This is a reason why physics departments performed comparatively better than those in two other disciplines.

9.2 Link Patterns along National and Disciplinary Lines

The link patterns identified along national lines are the answer to the hypothesis 'H2: Departmental link patterns differ along country lines.' Those patterns identified along disciplinary lines are the answer to the hypothesis 'H3: Departmental link patterns differ along disciplinary lines, even for similar disciplines.' The link patterns for the three types of departments in the three countries are identified with regard to the four aspects.

9.2.1 Link Patterns in Four Aspects

9.2.1.1 General Web Uses

- Departmental web site sizes

Along national lines, UK departments have the largest web sites, followed by Canada and then Australia. Along disciplinary lines, chemistry departments have the smallest web sites in each country, while physics departments have the largest.

- Departmental external inlinks or outlinks

Along national lines, UK departments attract the most external inlinks, followed by those in Canada and then Australia; UK departments create the most outlinks, followed by those in Canada and then Australia. Along disciplinary lines, physics departments attract the most external inlinks, chemistry attract the least; physics departments create the most outlinks while chemistry departments create the least.

- Proportion of national inlinks

Along national lines, UK departments have the smallest proportion of national inlinks, followed by those in Canada and then Australia. Along disciplinary lines, proportion of national inlinks for physics departments is the smallest, while that of chemistry departments is the largest.

9.2.1.2 National Peer Interlinks

- Interconnection rate

Along national lines, Canadian departments' interconnection rate is the largest, followed by Australia and then the UK. Along disciplinary lines, physics departments' interconnection rate is the largest, followed by chemistry and biology.

- National peer inlinks

Along national lines, Canadian depargtments' have the largest median peer inlink counts, followed by Australia and then the UK. Along disciplinary lines, physics departments have the largest median national peer inlink counts while biology departments have the smallest median national peer inlink coutns.

9.2.1.3 International Peer Interlinks

- Link Propensities

Along national lines, the LP for UK departments is the largest, followed by Australia and then Canada. Along disciplinary lines, physics departments' link propensity is the largest while that of biology departments is the smallest.

- Adapted mean international peer inlinks

Along national lines, the UK departments receive the largest adapted mean international peer inlinks, followed by Australia and then Canada. Along disciplinary lines, physics departments receive the largest adapted mean international peer inlinks while biology departments receive the smallest.

9.2.1.4 Interactions with Different Top Level Domains

Along national lines, Canadian departments attract the most edu inlinks while those from the UK attract the smallest edu inlinks; Canadian departments outlink to edu domain the most amongst the three countrie; Australian departments attract the smallest com inlinks amongst the three countries, however, they outlink the most to the com domain; with regard to the country domains, each set of departments outlink to its own country domains the most, and then Canadian and Australian departments tend to link to the uk domain, while UK departments tend to link to the au domain.

Along disciplinary lines, physics departments outlink to the edu domain the most; chemistry departments outlink to the org, edu and net domains the least, while they outlink to the com and uk domains the most.

9.2.2 Summary of General Link Patterns

The number of web pages and outlinks are reflections of a set of departments' web publishing activity level. The number of inlinks shows the influence of a set of departments, with peer inlinks showing the influence over peers and external inlinks showing the influence over the Web.

9.2.2.1 Link Patterns along National Lines

- The UK departments make the most general use of the Web, followed by Canada and then Australia.
- The Canadian departments interconnect the best with their national peers, followed by Australia and then the UK.
- Departments in the UK attract the most international peer inlinks, followed by Australia and then Canada.
- With regard to au, ca and uk domains, each set of departments link the most to its own country domain, then the departments from both Australia and Canada link to the uk domain. The UK departments link more to the au than ca domain.
- Among the three countries, the Canadian departments outlink to the edu domain the most, while the Australian departments outlink to the com domain the most.

The UK departments make the most general use of the Web. Canadian departments interconnect the best with their national peers, however, the worst with their international peers. The influence of Canadian departments over their international peers may be biased by the limited selection of countries in this study. The above results illustrate that Australian and UK departments have a relatively closer relationship on the Web. This is a reflection of a closer collaboration in formal scholarly communication between the two countries. The fact that Canadian departments link the most to the edu domain reflects the closer relationship between Canada and the United States.

9.2.2.2 Link Patterns along Disciplinary Lines

- Physics departments make the most general use of the Web, followed by Biology and then Chemistry.
- Physics departments interlink the most with their national and international peers, followed by Chemistry and then Biology.
- Physics departments outlink to the edu domain the most.

Physics departments make the most general use of the Web, while chemistry the least; for influences over both national and international peers, physics perform the best and biology the least. This may be a reflection of physics being the hardest science out of the three disciplines and making the most use of the Web. Chemistry and biology are similar, comparing with physics, and patterns are different for them with regard to general web use and peer interlinking. This may also be a reflection of different ways for researchers from different disciplines to deal with their research outputs, with physicists being more open, and chemists and biologists being more conservative.

9.2.2.3 Significance of the Link Patterns Identified

Different link patterns have been identified along national and disciplinary lines. This is very useful for future investigations into departmental web use in webometrics. As departments from same disciplines in the three similar countries employ the Web differently, departments from more different countries may well employ the Web differently to an even great extent. Similarly, although the three disciplines in the project are similar, different link patterns have been identified and the patterns are consistent in the three countries. This is in parallel with findings from citation analysis that various disciplines use citations differently. Just as it is not relevant to compare the research performance of departments from different disciplines according to their citation counts, it is also not relevant to compare web performance between different types of departments within one university according to their link counts.

Another interesting finding is that the link patterns reflect off-line phenomena. Departments with better research performances tend to make more web publications and create more links. With regard to international peer interlinking, countries having more formal scholarly communications tend to interlink more with each other. Departments from a discipline that are more willing to distribute their formal research outputs also tend to create more web pages and links (elsewhere or to their peers). Even among similar departments, harder disciplines tend to make more use of the Web.

9.3 Future Work

Future research can go in two directions. One is to study on more disciplines in one country; another is to study one discipline in more countries. The results from the first approach may illustrate different link patterns along disciplinary lines, while the results from the second approach may show different link patterns for one discipline along national lines.

The three disciplines in this project are similar. If more disciplines are added, for example, in addition to physics, chemistry and biology, departments in law, mathematics, engineering, linguistic, business and computing are included, a fuller picture of different link patterns along disciplinary lines can be identified for a country.

The Canadian departments' ability to attract international peer inlinks is biased by considering international peer inlinks from only the UK and Australia. If the United States were included in this study, the Canadian departments' ability to attract international peer inlinks would have been better. It is sensible to choose one discipline, such as physics, and include more countries in future research, for example, the United States and Western European countries, to seek international interlinking patterns. Hopefully the patterns can be identified more accurately, and nations that do not exploit the Web efficiently can also be notified of the need for remedial action.

References

Acosta, M. & Coronado, D. (2003) Science-technology flows in Spanish regions: An analysis of scientific citations in patents, *Research Policy,* 32(10), 1783-1803.

Aksnes, D. W. & Taxt, R. E. (2004) Peer reviews and bibliometrics indicators: A comparative study at a Norwegian university, *Research Evaluation,* 13(1), 33-41.

AllTheWeb (2002) Available at: http://www.alltheweb.com/ (12 Feb. 2002)

Almind, T. C. & Ingwersen, P. (1997) Informetric analyses on the World Wide Web: Methodological approaches to 'webometrics', *Journal of Documentation,* 53(4), 404-426.

AltaVista (2001) Available at: http://www.altavista.com/ (12 Oct. 2001)

AsiaWeek (2000) Asia's best universities 2000, Available at: http://www.asiaweek.com/asiaweek/features/universities2000/rank_country/ (12 Feb. 2003)

Baird, L. M. & Oppenheim, C. (1994) Do citations matter?, *Journal of Information Science,* 20(1), 2-15.

Baldi, S. (1998) Normative versus social constructivist processes in the allocation of citations: A network-analytic model, *American Sociological Review,* 63(6), 829-846.

Balnaves, M. & Caputi, P. (2001) Introduction to quantitative research methods: An investigative approach, London: Sage Publications.

Bar-Ilan, J. (1999) Search engine results over time: A case study on search engine stability, *Cybermetrics,* 2/3(1). Available at: http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html

Bar-Ilan, J. (2001) Data collection methods on the Web for informetric purposes: A review and analysis, *Scientometrics,* 50(1), 7-32.

Bar-Ilan, J. (2004a) A microscopic link analysis of academic institutions within a country: The case of Israel, *Scientometrics,* 59(3), 391-403.

Bar-Ilan, J. (2004b) Self-linking and self-linked rates of academic institutions on the Web, *Scientometrics,* 59(1), 29-41.

Bar-Ilan, J. (2005) What do we know about links and linking? A framework for studying links in academic environments, *Information Processing and Management,* 41(4), 973-986.

Bayer, A. E., Smart, J. C. & McLaughlin, G. W. (1990) Mapping intellectual structure of a scientific subfield through author cocitations, *Journal of the American Society for Information Science,* 41(6), 444-452.

Bence, V. & Oppenheim, C. (2004a) The influence of peer review on the research assessment exercise, *Journal of Information Science,* 30(4), 347-368.

Bence, V. & Oppenheim, C. (2004b) The role of academic journal publications in the UK research assessment exercise, *Learned Publishing,* 17(1), 53-68.

Berners-Lee, T., Fielding, R. & Masinter, L. (1998) Uniform resource identifiers (URI): Generic syntax, Available at: http://www.ietf.org/rfc/rfc2396.txt (4 Sept. 2004)

Bessant, J., Birley, S., Cooper, C., Dawson, S., Gennard, J., Gardiner, M., Gray, A., Jones, P., Mayer, C., McGee, J., Pidd, M., Rowley, G., Saunders, J. A. & Stark, A. (2003) The state of the field in UK management research:

Reflections of the research assessment exercise (RAE) panel, *British Journal of Management,* 14(1), 51-68.
Bharat, K. & Broder, A. (1998) A technique for measuring the relative size and overlap of public web search engines, *Computer Networks and ISDN Systems,* 30(1-7), 379-388. Available at: http://www7.scu.edu.au/programme/fullpapers/1937/com1937.htm#bio2 (12 Feb. 2002)
Björneborn, L. (2001a) Shared outlinks in small-world co-linkage analysis: A webometric pilot study of bibliographic couplings on researchers' bookmark lists on the Web, Royal School of Library and Information Science, Denmark.
Björneborn, L. (2001b) Small-world linkage and co-linkage, In *Proceeding of the 12th ACM Conference on Hypertext and Hypermedia*, Århus, Denmark, pp 133-134.
Björneborn, L. (2004) Small-world link structures across an academic web space: A library and information science approach, Denmark: Royal School of Library and Information Science.
Björneborn, L. & Ingwersen, P. (2001) Perspectives of webometrics, *Scientometrics,* 50(1), 65-82.
Björneborn, L. & Ingwersen, P. (2004) Toward a basic framework for webometrics, *Journal of the American Society for Information Science and Technology,* 55(14), 1216-1227.
Black, T. R. (1999) Doing quantitative research in the social sciences: An integrated approach to research design, measurement and statistics, London: Sage Publications.
Bonzi, S. & Snyder, H. W. (1991) Motivations for citation: A comparison of self citation and citation to others, *Scientometrics,* 21(2), 245-254.
Borgman, C. L. & Furner, J. (2002) Scholarly communication and bibliometrics, In: Cronin, B. (ed), *Annual Review of Information Science and Technology*, Vol. 36, Medford, NY: Information Today Inc., pp. 3-72.
Bourke, P. & Butler, L. (1998) Institutions and the map of science: Matching university departments and fields of research, *Research Policy,* 26(6), 711-718.
Bradford, S. C. (1934) Sources of information on specific subjects, *British Journal of Engineering,* 137, 85-86.
Brin, S. & Page, L. (1998) The anatomy of a large-scale hypertextual web search engine, In *Proceeding of 7th International WWW Conference*, Brisbane, Australia, 14-18 Apr., pp 107-117.
Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. & Wiener, J. (2000) Graph structure in the Web, *Computer Networks - The International Journal of Computer and Telecommunications Networking,* 33(1-6), 309-320.
Brooks, T. A. (1985) Private acts and public objects: An investigation of citer motivations, *Journal of the American Society for Information Science,* 36(4), 223-229.
Brooks, T. A. (1986) Evidence of complex citer motivations, *Journal of the American Society for Information Science,* 37(1), 34-36.
Butler, L. (2003) Explaining Australia's increased share of ISI publications: The effects of a funding formula based on publication counts, *Research Policy,* 32(1), 143-155.

Cano, V. (1989) Citation behavior: Classification, utility, and location, *Journal of the American Society for Information Science*, 40(4), 284-290.
Carpenter, M. P., Cooper, M. & Narin, F. (1980) Linkage between basic research literature and patents, *Research Management*, 23(2), 30-35.
Carpenter, M. P. & Narin, F. (1983) Assessment of the linkages between patents and fundamental research, *Scientometrics*, 5(1), 78-79.
Case, D. O. & Higgins, G. M. (2000) How can we investigate citation behaviour? A study of reasons for citing literature in communication, *Journal of the American Society for Information Science*, 51(7), 635-645.
Chen, C., Newman, J., Newman, R. & Rada, R. (1998) How did university departments interweave the Web: A study of connectivity and underlying factors, *Interacting with Computers*, 10(4), 353-373.
Chu, H. (2005) Taxonomy of inlinked web entities: What does it imply for webometric research?, *Library and Information Science Research*, 27(1), 8-27.
Chu, H., He, S. & Thelwall, M. (2002) Library and information science schools in Canada and USA: A webometric perspective, *Journal of Education for Library and Information Science*, 43(2), 110-125.
Chubin, D. E. & Moitra, S. D. (1975) Content analysis of references: Adjunct or alternative to citation counting?, *Social Studies of Science*, 5(4), 423-441.
CIA (2001) World Factbook 2000, Available at: http://www.odci.gov/cia/publications/factbook/index.html (4 May 2001)
Cole, J. R. & Cole, S. (1967) Scientific output and recognition: A study in the operation of the reward system in science, *American Sociological Review*, 32(3), 377-390.
Cole, J. R. & Cole, S. (1973) Social stratification in science, Chicago: University of Chicago Press
Coronado, D., Acosta, M. & Leon, D. (2004) Regional planning of R&D and science-technology interactions in Audalucia: A bibliometric analysis of patent documents, *European Planning Studies*, 12(8), 1075-1095.
Coward, H. R. & Franklin, J. J. (1989) Identifying the science-technology interface: Matching patent data to a bibliometric model, *Science Technology & Human Values*, 14(1), 50-77.
Cronin, B. (1984) The citation process: The role and significance of citations in scientific communication, London: Taylor Graham
Cronin, B. (2001) Bibliometrics and beyond: Some thoughts on web-based citation analysis, *Journal of Information Science*, 27(1), 1-7.
Cronin, B., Mckenzie, G., Rubio, L. & Weaverwozniak, S. (1993) Accounting for influence: Acknowledgements in contemporary sociology, *Journal of the American Society for Information Science*, 44(7), 406-412.
Cronin, B. & Overfelt, K. (1994) Citation-based auditing of academic-performance, *Journal of American Society for Information Science*, 45(2), 61-72.
Cronin, B., Snyder, H. W., Rosenbaum, H., Martinson, A. & Callahan, E. (1998) Invoked on the web, *Journal of the American Society for Information Science*, 49(14), 1319-1328.
DDC (2003) Dewey decimal classification, Available at: http://www.oclc.org/dewey/ (1 Oct. 2003)

DEST (2005) Institutional grants scheme, Available at: http://www.dest.gov.au/archive/highered/he_report/2002_2004/html/3_3.htm (11 May 2005)
DETYA (2001) Higher education report for the 2000 to 2002 triennium, Available at: http://www.detya.gov.au/highered/he_report/2000_2002/html/3_8.htm (1 Mar. 2003)
Edge, D. (1979) Quantitative measures of communication in science: A critical review, *History of Science,* 17(36), 102-134.
EducationGuardian (2001) Institutional tables, Available at: http://education.guardian.co.uk/researchratings/0,11227,616612,00.html (18 Mar. 2002)
Egghe, L. (2000) New informetric aspects of the internet: Some reflections - many problems, *Journal of Information Science,* 26(5), 329-335.
Egghe, L. (2001) Discussions on informetrics of the internet, In *Proceeding of the 2nd International Symposium on Quantitative Research Evaluation and the 6th National Annual Conference of Scientometrics and Informetrics*, Shanghai, pp 8-26.
Egghe, L. & Rousseau, R. (1990) Introduction to informetrics: Quantitative methods in library, documentation and information science, Amsterdam: Elsevier
Ellis, P., Hepburn, G. & Oppenheim, C. (1978) Studies on patent citation networks, *Journal of Documentation,* 34(1), 12-20.
Garfield, E. (1955) Citation indexes for science: A new dimension in documentation through association of ideas, *Science,* 122(3159), 108-111. Available at: http://garfield.library.upenn.edu/essays/v6p468y1983.pdf
Garfield, E. (1965) Can citation indexing be automated?, In: Heilprin, L. (ed), *Statistical association methods for mechanized documentation: Symposium proceedings* Washington, DC: National Bureau of Standards, pp. 189-192.
Garfield, E. (1975) The 'obliteration phenomenon' in science and the advantage of being obliterated, *Current Contents,* 51/52, 5-7.
Garfield, E. (1979) Is citation analysis a legitimate evaluation tool?, *Scientometrics,* 1(4), 359-375.
Garfield, E. (1987) Lauching the ISI Atlas of Science: For the new year, a new generation reviews, *Current Contents,* 1, 3-8.
Garfield, E. (1996) When to cite, *Library Quarterly,* 66(4), 449-458.
Garfield, E. (1999) Journal impact factor: A brief review, *Canadian Medical Association Journal,* 161(8), 979-980. Available at: http://www.garfield.library.upenn.edu/papers/journalimpactCMAJ1999.pdf
Garner, R. (1967) A computer oriented, graph theoretic analysis of citation index structures, In: Flood, B. (ed), *Three Drexel Information Science Research Studies* Drexel Press, pp. 3-46.
Geuna, A. & Martin, B. R. (2003) University research evaluation and funding: An international comparison, *Minerva,* 41(4), 277-304.
Gilbert, G. N. (1977) Referencing as persuasion, *Social Studies of Science,* 7(1), 113-122.

Gillett, R. (1991) Pitfalls in assessing research performance by grant income, *Scientometrics*, 22(2), 253-263.
Glänzel, W. (2001) National characteristics in international scientific co-authorship relations, *Scientometrics*, 51(1), 69-115.
Glänzel, W. (2003) Bibliometrics as a research field: A course on theory and application of bibliometric indicators, Available at: http://www.norslis.net/2004/Bib_Module_KUL.pdf (25 Apr. 2005)
Glänzel, W. & Meyer, M. (2003) Patents cited in the scientific literature: An exploratory study of 'reverse' citation relations, *Scientometrics*, 58(2), 415-428.
Glänzel, W. & Schoepflin, U. (1999) A bibliometric study of reference literature in the sciences and social sciences, *Information Processing and Management*, 35(1), 31-44.
Glänzel, W. & Schubert, A. (2001) Double effort = Double impact? A critical view at international co-authorship in chemistry, *Scientometrics*, 50(2), 199-214.
Glänzel, W. & Thus, B. (2004) The influence of author self-citations on bibliometric macro indicators, *Scientometrics*, 59(3), 281-310.
Google (2002) Our search: Google technology, Available at: http://www.google.com/technology/ (5 Feb. 2003)
Grupp, H. & Schnoring, T. (1992) Research-and-development in telecommunications: National systems under pressure, *Telecommunications Policy*, 16(1), 46-66.
Haas, S. W. & Grams, E. S. (2000) Readers, authors and page structure: A discussion of four questions arising from a content analysis of web pages, *Journal of American Society for Information Science*, 51(2), 181-192.
Harries, G., Wilkinson, D., Price, E., Fairclough, R. & Thelwall, M. (2004) Hyperlinks as a data source for science mapping, *Journal of Information Science*, 30(5), 436-447.
HEFCE (2005) Research Assessment Exercise, Available at: http://www.hefce.ac.uk/research/assessment/ (16 May 2005)
Henderson, R., Jaffe, A. B. & Trajtenberg, M. (1998) Universities as a source of commercial technology: A detailed analysis of university patenting, 1965-1988, *Review of Economics and Statistics*, 80(1), 119-127.
Herman, I. L. (1991) Receptivity to foreign literature: A comparison of UK and United States citating behavior in librarianship and information science, *Library and Information Science Research*, 13(1), 37-47.
HERO (2001a) 3 Description of the RAE results, Available at: http://www.hero.ac.uk/rae/Pubs/4_01/section3.htm (20 Feb. 2002)
HERO (2001b) 4 Units of Assessment, Available at: http://www.hero.ac.uk/rae/Pubs/4_01/section4.htm (20 Feb. 2002)
HERO (2001c) A guide to the 2001 Research Assessment Exercise, Available at: http://www.hero.ac.uk/rae/Pubs/other/raeguide.pdf (18 Mar. 2002)
HERO (2001d) RAE 2001, Available at: http://www.hero.ac.uk/rae/ (1 Mar. 2002)
HERO (2001e) RAE 2001 Submissions, Available at: http://www.hero.ac.uk/rae/submissions/ (12 Jan. 2005)
HESA (2002) Staff FTEs by institution and cost centre 1998/99, Available at: http://www.hesa.ac.uk/acuk/maninfo/1998-99/staf_fte/stafftc89.htm (23 Jan. 2002)

Holmes, A. & Oppenheim, C. (2001) Use of citation analysis to predict the outcome of the 2001 Research Assessment Exercise for Unit of Assessment (UoA) 61: Library and Information Management, *Information Research,* 6(2). Available at: http://informationr.net/ir/6-2/paper103a.html (12 Feb. 2002)
Hornbostel, S. (2001) Third party funding of German universities. An indicator of research activity?, *Scientometrics,* 50(3), 523-537.
Hummon, N. P. & Doreian, P. (1989) Connectivity in a citation network: The development of DNA theory, *Social Networks,* 11(1), 39-63. Available at: http://www.garfield.library.upenn.edu/papers/hummondoreian1989.pdf
IAU (2003a) International handbook of universities, London: Palgrave
IAU (2003b) World higher education database 2003/4 CD-Rom, London: Palgrave
ICANN/GNSO (2001) gTLD registries, Available at: http://www.dnso.org/constituency/gtld/gtld.html (13 Feb. 2002)
Ingwersen, P. (1998) The calculation of web impact factors, *Journal of Documentation,* 54(2), 236-243.
ISI (2001a) Glossary, Available at: http://portalt.wok.mimas.ac.uk/portal.cgi?DestApp=JCR&Func=Frame (1 Mar. 2001)
ISI (2001b) ISI Web of Knowledge, Available at: http://wok.mimas.ac.uk/ (15 Oct. 2001)
ISI (2003) Rankings for the 20 most-cited countries in all fields, Available at: http://www.in-cites.com/countries/2003allfields.html (9 June 2004)
ISI (2005) ISI web of science, Available at: http://www.isinet.com/isi/products/citation/wos/ (12 May 2005)
Iversen, E. J. (2000) An excursion into the patent-bibliometrics of Norwegian patenting, *Scientometrics,* 49(1), 63-80.
Johnes, J. & Johnes, G. (1995) Research funding and performance in UK university departments of economics: A frontier analysis, *Economics of Education Review,* 14(3), 301-314.
Kaback, S. M., Lambert, N. & Simmons, E. S. (1994) Patent citation data: Source of insights, or nothing to get excited about, *Abstracts of Papers of the American Chemical Society,* 208(44), Part 1.
Katz, J. S. & Hicks, D. (1997) How much is a collaboration worth? A calibrated bibliometric model, *Scientometrics,* 40(3), 541-554.
Kessler, M. M. (1963) Bibliographic coupling between scientific papers, *American Documentation,* 14(1), 10-25.
Kim, H. J. (2000) Motivations for hyperlinking in scholarly electronic articles: A qualitative study, *Journal of the American Society for Information Science,* 51(10), 887-899.
Kleinberg, J. M. (1999) Authoritative sources in a hyperlinked environment, *Journal of the ACM,* 46(5), 604-632.
Kling, R. & McKim, G. (2000) Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication, *Journal of the American Society for Information Science,* 51(14), 1306-1320.
Koehler, W. (2002) Web page change and persistence: A four-year longitudinal study, *Journal of American Society for Information Science,* 53(2), 162-171.

Lange, L. L. (2001) Citation counts of multi-authored papers: First-named authors and further authors, *Scientometrics,* 52(3), 457-470.
Larson, R. R. (1996) Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace, In *Proceeding of the ASIS 59th annual meeting*, pp 71-78. Available at: http://sherlock.berkeley.edu/asis96/asis96.html
Laudel, G. (2002) What do we measure by co-authorships?, *Research Evaluation,* 11(1), 3-15.
Lawrence, S., Coetzee, F. M., Glover, E. J., Pennock, D., Flake, G. W., Nielsen, F., Krovetz, B., Kruger, A. & Giles, C. L. (2001) Persistence of web references in scientific research, *IEEE Computer,* 34(2), 26-31.
Lawrence, S. & Giles, C. L. (1999) Accessibility of information on the web, *Nature,* 400(6740), 107-109. Available at: http://wwwmetrics.com/
Lawrence, S. L. (2001) Free online availability substantially increases a paper's impact, *Nature,* 411(6837), 521.
LCB (2003) Library of congress bibliographies, Available at: http://www.loc.gov/rr/bibguide.html (1 Oct. 2003)
Leydesdorff, L. (2004) Clusters and maps of science journals based on bi-connected graphs in Journal Citation Reports, *Journal of Documentation,* 60(4), 371-427.
Leydesdorff, L. & Curran, M. (2000) Mapping university-industry-government relations on the Internet: the construction of indicators for a knowledge-based economy, *Cybermetrics,* 4(1), Paper 2. Available at: http://www.cindoc.csic.es/cybermetrics/articles/v4i1p2.html
Leydesdorff, L. & Wouters, P. (1999) Between texts and contexts: Advances in theories of citation?, *Scientometrics,* 44(2), 169-182.
Li, X. (2003) A review of the development and application of the web impact factor, *Online Information Review,* 27(6), 407-417.
Li, X., Thelwall, M., Musgrove, P. & Wilkinson, D. (2003a) Do academic web pages in western europe attract more links if they are in English?, In *Proceeding of 9th ISSI*, Beijing, China, 25-29 August 2003,
Li, X., Thelwall, M., Musgrove, P. & Wilkinson, D. (2003b) The relationship between the WIFs or inlinks of computer science departments in UK and their RAE ratings or research productivities in 2001, *Scientometrics,* 57(2), 239-255.
Lindsey, D. (1980) Production and citation measures in the sociology of science: The problem of multiple authorship, *Social Studies of Science,* 10(2), 145-162.
Liu, M. (1993) Progress in documentation - The complexities of citation practice - A review of citation studies, *Journal of Documentation,* 49(4), 370-408.
Long, J. S. & McGinnis, R. (1982) On adjusting productivity measures for mutiple authorship, *Scientometrics,* 4(5), 379-387.
Lotka, A. J. (1926) The frequency distribution of scientific productivity, *Journal of the Washington Academy of Sciences,* 16(12), 317-323.
Luwel, M., Moed, H. F., Nederhof, A. J., De Samblanx, V., Verbrugghen, K. & van der Wurff, L. J. (1999) Towards indicators of research performance in the social sciences and humanities, Leiden: CWTS. Available at: http://www.cwts.nl/cwts/2wrds_indi_vl_99.html

MacRoberts, M. H. & MacRoberts, B. R. (1989) Problems of citation analysis - A critical review, *Journal of the American Society for Information Science,* 40(5), 342-349.
Markwell, J. & Brooks, D. W. (2002) Broken links: The ephemeral nature of educational WWW hyperlinks, *Journal of Science Education and Technology,* 11(2), 105-108.
Martin, B. R. (1996) The use of multiple indicators in the assessment of basic research, *Scientometrics,* 36(3), 343-362.
Mayfield University Consultants (2000) League tables 2000, *The Times Higher Education Supplement,* April 14, II-III.
McCain, K. W. (1986) Cocited author mapping as a valid representation of intellectual structure, *Journal of American Society for Information Science,* 37(3), 111-122.
McCain, K. W. (1991) Mapping economics through the journal literature: An experiment in journal cocitation analysis, *Journal of the American Society for Information Science,* 42(4), 290-296.
McKiernan, G. (1996) CitedSites(sm): citation indexing of web resources, Available at: http://www.public.iastate.edu/~CYBERSTACKS/Cited.htm (12 Feb. 2002)
McMillan, G. S., Narin, F. & Deeds, D. L. (2000) An analysis of the critical role of public science in innovation: The case of biotechnology, *Research Policy,* 29(1), 1-8.
Medcalc (2003) Medcalc, Available at: http://www.medcalc.be/manual/mpage08-06.php (18 May 2003)
Meghabghab, G. (2002) Discovering authorities and hubs in different topological web graph structures, *Information Processing and Management,* 38(1), 111-140.
Melin, G. & Persson, O. (1996) Studying research collaboration using co-authorships, *Scientometrics,* 36(3), 363-377.
Merton, R. K. (1968a) The Matthew effect in science, *Science,* 159(3810), 56-63.
Merton, R. K. (1968b) Social theory and social structure, New York: Free Press
Merton, R. K. (1973) The sociology of science: Theoritical and empirical investigations, Chicago: University of Chicago Press
Mettrop, W. & Nieuwenhuysen, P. (2001) Internet search engines: Fluctuations in document accessibility, *Journal of Documentation,* 57(5), 623-651.
Meyer, M. (2000) Does science push technology? Patents citing scientific literature, *Research Policy,* 29(3), 409-434.
Meyer-Krahmer, F. & Schmoch, U. (1998) Science-based technologies: University-industry interactions in four fields, *Research Policy,* 27(8), 835-851.
Moed, H. (2005) Citation analysis in research evaluation, Dordrecht: Springer
Moed, H. F. (2002) The impact-factors debate: The ISI's uses and limits, *Nature,* 415(6873), 731-732.
Moed, H. F., Burger, W. J. M., Frankfort, J. G. & van Raan, A. F. J. (1985) A comparative study of bibliometric past performance analysis and peer judgement, *Scientometrics,* 8, 149-159.
Moed, H. F., de Bruin, R. E. & van Leeuwen, T. (1995) New bibliometric tools for the assessment of national research performance - Database description, overview of indicators and first applications, *Scientometrics,* 33(3), 381-422.

Moravcsik, M. J. (1973) Measures of scientific growth, *Research Policy,* 2(3), 266-275.

Moravcsik, M. J. & Murugesan, P. (1975) Some results on the function and quality of citations, *Social Studies of Science,* 5(1), 86-92.

Moxham, H. & Anderson, J. (1992) Peer review. A view from the inside, *Science and Technology Policy,* 5(1), 7-15.

Narin, F. (1994) Patent citation indicators, *Abstracts of Papers of the American Chemical Society,* 208(42), Part 1.

Narin, F., Breitzman, T. & Thomas, P. (2004) Using patent citation indicators to manage a stock portfolio, In: Moed, H., Glänzel, W. and Schmoch, U. (eds), *Handbook of Quantitative Social Science and Technology Research* Dordrecht: Kluwer Academic Publishers, pp. 553-568.

Narin, F., Hamilton, K. S. & Olivastro, D. (1997) The increasing linkage between U.S. technology and public science, *Research Policy,* 26(3), 317-330.

Narin, F. & Noma, E. (1985) Is technology becoming a science?, *Scientometrics,* 7(3-6), 369-381.

Narin, F. & Olivastro, D. (1992) Linkage between technology and science, *Research Policy,* 21(3), 237-249.

Narin, F. & Olivastro, D. (1998) Linkage between patents and papers: An interim EPO/US comparison, *Scientometrics,* 41(1-2), 51-59.

NetBig (2001) Chinese university ranking 2001, Available at: http://rank2001.netbig.com/en/about/03.htm (1 Apr. 2003)

Norris, M. & Oppenheim, C. (2003) Citation counts and the research assessment exercise V: Archaeology and the 2001 RAE, *Journal of Documentation,* 59(6), 709-730.

Novak, J. D. & Gowin, D. B. (1984) Learning how to learn, Cambridge: Cambridge University Press

Noyons, E. C. M., van Raan, A. F. J., Grupp, H. & Schmoch, U. (1994) Exploring the science and technology interface: Inventor author relations in laser medicine research, *Research Policy,* 23(4), 443-457.

NSERC (2003) 2003 research grants competition results by GSC, Available at: http://www.nserc.ca/programs/result/2003/rg/gsc_e.htm (20 Nov. 2003)

Oppenheim, C. (1995) The correlation between citation counts and the 1992 research assessment exercise ratings for British library and information science university departments, *Journal of Documentation,* 51(1), 18-27.

Oppenheim, C. (1997) The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology, *Journal of Documentation,* 53(5), 477-487.

Oppenheim, C. (2000) Do patent citations count?, In: Cronin, B. and Atkins, H. B. (eds), *The web of knowledge: a festschrift in honor of Eugene Garfield* Metford, NJ: Information Today Inc. ASIS Monograph Series, pp. 405-432.

Oppenheim, C. & Renn, S. (1978) Highly cited old papers and the reasons why they continue to be cited, *Journal of the American Society for Information Science,* 29(5), 225-231.

Oppenheim, C. & Smith, R. (2001) Student citation practices in an information science department, *Education for Information,* 19(4), 299-323.

Pandia (2004) The death of AltaVista and AlltheWeb, Available at: http://www.pandia.com/sw-2004/08-yahoo.html (1 Apr. 2004)

Park, H. W. & Thelwall, M. (2006, to appear) Web science communicaton in the age of globalization: Links among universities' websites in Asia and Europe, *New Media and Society*.
Pennock, D., Flake, G. W., Lawrence, S., Glover, E. J. & Giles, C. L. (2002) Winners don't take all: Characterizing the competition for links on the web, *Proceedings of the National Academy of Sciences of the United States of America*, 99(8), 5207-5211.
Peritz, B. C. (1983) A classification of citation roles for the social sciences and related fields, *Scientometrics*, 5(5), 303-312.
Phillimore, A. J. (1989) University research performance indicators in practice - The university grants committees evaluation of British-universities, 1985-86, *Research Policy*, 18(5), 255-271.
Pinski, G. & Narin, F. (1976) Citation influence for journal aggregates of scientific publications: Theory with application to the literature of physics, *Information Processing and Management*, 12(5), 297-312.
Price, E. & Thelwall, M. (2005, to appear) The clustering power of low frequency words in academic webs, *Journal of the American Society for Information Science and Technology*.
Prime-Claverie, C., Bassecoulard, E. & Zitt, M. (2002) Co-citations and co-sitations: A cautionary view on an analogy, *Scientometrics*, 54(2), 291-308.
Prime-Claverie, C. & Beigbeder, M. (2004) Transposition of the cocitation method with a view to classifying web pages, *Journal of the American Society for Information Science and Technology*, 55(14), 1282-1289.
Ren, S. & Zu, G. (2002) Statistics hide impact of non-English journals, *Nature*, 415(6873), 732.
REPP (2003) Research evaluation and policy project, Available at: http://repp.anu.edu.au/ (21 Sept. 2003)
Rinia, E. J., van Leeuwen, T., van Vuren, H. G. & van Raan, A. F. J. (1998) Comparative analysis of a set of bibliometric indicators and central peer review criteria: Evaluation of condensed matter physics in the Netherlands., *Research Policy*, 27(1), 95-107.
Rodríguez i Gairín, J. M. (1997) Valorando el impacto de la información en Internet: Altavista, el "Citation Index" de la Red [Impact assessment of information on the Internet: AltaVista, the citation index of the web], *Revista Espanola de Documentacion Scientifica*, 20(2), 175-181. Available at: http://www.kronosdoc.com/publicacions/altavis.htm
Rousseau, R. (1992) Why am I not cited or why are multiauthored papers more cited than others, *Journal of Documentation*, 48(1), 79-80.
Rousseau, R. (1997) Sitations: an exploratory study, *Cybermetrics*, 1(1). Available at: http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html (12 Nov. 2001)
Rousseau, R. (1999) Daily time series of common single word searches in AltaVista and NorthernLight, *Cybermetrics*, 2/3. Available at: http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html
Schmoch, U. (1993) Tracing the knowledge transfer from science to technology as reflected in patent indicators, *Scientometrics*, 26(1), 193-211.
Seglen, P. O. (1998) Citation rates and journal impact factors are not suitable for evaluation of research, *ACTA Orthopaedica Scandinavica*, 69(3), 224-229.

Shadish, W. R., Tolliver, D., Gray, M. & Sengupta, S. K. (1995) Author judgments about works they cite: 3 studies from psychology journals, *Social Studies of Science*, 25(3), 477-498.
Sharabchiev, J. T. (1988) Comparative analysis of two methods of cluster analysis of bibliographic references, *Nauchno-Teknicheskaya Informatsiya*, 2(4), 25-28.
Simmons, E. S. (1994) Getting from here to there: Shortcuts for patent searchers, *Abstracts of Papers of the American Chemical Society*, 207(29), part 1.
Small, H. (1973) Co-citation in the scientific literature: A new measure of the relationship between two documents, *Journal of the American Society for Information Science*, 24(4), 265-269.
Small, H. (1983) Citation context analysis, In: Voigt, M. (ed), *Progress in communication science*, Vol. 3, Norwood, NJ: Ablex, pp. 287-310.
Small, H. (1998) A general framework for general large-scale maps of science in two or three dimensions: The SciViz system, *Scientometrics*, 41(1-2), 125-133.
Small, H. (1999a) A passage through science: Crossing disciplinary boundaries, *Library Trends*, 48(1), 72-108.
Small, H. (1999b) Visualizing science through citation mapping, *Journal of the American Society for Information Science*, 50(9), 799-813.
Small, H. (2000) Charting pathways through science: Exploring Garfield's vision of a unified index to science, In: Cronin, B. and Atkins, H. B. (eds), *The web of knowledge: A festschrift in honor of Eugene Garfield* Medford, NJ: Information Today, Inc. ASIS Monograph Series, pp. 449-473.
Smith, A. & Eysenck, M. (2002) The correlation between RAE ratings and citation counts in psychology, University of London. Available at: http://cogprints.ecs.soton.ac.uk/archive/00002749/01/citations.pdf (15 Jan. 2005)
Smith, A. G. (1999a) ANZAC Webometrics: Exploring Australasian web structures, In *Proceeding of Information Online and On Disc 99*, Sydney, Australia, 19-21 January 1999, pp 159-181. Available at: http://www.csu.edu.au/special/online99/proceedings99/203b.htm.
Smith, A. G. (1999b) The impact of web sites: A comparison between Australasia and Latin America, In *Proceeding of INFO'99, Congreso Internacional de Informacion*, Havana, 4-8 October 1999, Available at: http://www.vuw.ac.nz/~agsmith/publns/austlat/
Smith, A. G. (1999c) A tale of two web spaces: Comparing sites using web impact factors, *Journal of Documentation*, 55(5), 577-592.
Smith, A. G. (2004) Web links as analogues of citations, *Information Research*, 9(4). Available at: http://informationr.net/ir/9-4/paper188.html (12 Jan. 2005)
Smith, A. G. & Thelwall, M. (2001) Web impact factors and university research links, In *Proceeding of the 8th international Conference on Scientometrics & Informetrics*, Sydney, Australia, 16-20 July 2001, Vol.2, pp 657-664.
Smith, A. G. & Thelwall, M. (2002) Web impact factors for Australasian universities, *Scientometrics*, 54(3), 363-380.
Smith, L. (1981) Citation analysis, *Library Trends*, 30(1), 83-106.
Snyder, H. & Bonzi, S. (1998) Patterns of self-citation across disciplines (1980-1989), *Journal of Information Science*, 24(6), 431-435.

Snyder, H. W. & Rosenbaum, H. (1999) Can search engines be used as tools for web-link analysis? A critical review, *Journal of Documentation*, 55(4), 375-384.

Sullivan, D. (2001a) Search engine features, SearchEngineWatch, Available at: http://searchenginewatch.com/facts/assistance.html (28 Feb. 2003)

Sullivan, D. (2001b) Search engine sizes, SearchEngineWatch, Available at: http://searchenginewatch.com/reports/sizes.html (28 Feb. 2003)

Sullivan, D. (2002) Google tops in "Search Hours" ratings, SearchEngineWatch, Available at: http://searchenginewatch.com/sereport/02/05-ratings.html (6 Mar. 2003)

Tague-Sutcliffe, J. (1992) An introduction to informetrics, *Information Processing and Management*, 28(1), 1-3.

Tang, R. & Thelwall, M. (2002) Exploring the pattern of links between Chinese university web sites, In *Proceeding of the 65th Annual Meeting of American Society for Information Science and Technology*, 2002, Vol.39, pp 417-424.

Tang, R. & Thelwall, M. (2003a) Disciplinary differences in US academic departmental web site interlinking, *Library and Information Science Research*, 25(4), 437-458.

Tang, R. & Thelwall, M. (2003b) Patterns of international and national web inlinks to US university departments: A webometric analysis of disciplinary specificity, In *Proceeding of 9th ISSI conference*, Beijing, China, 25-29 Aug. 2003, pp 312-320.

Tang, R. & Thelwall, M. (2003c) US academic departmental web-site interlinking in the United States disciplinary differences, *Library and Information Science Research*, 25(4), 437-458.

Tang, R. & Thelwall, M. (2004) Patterns of national and international Web inlinks to US academic departments: An analysis of disciplinary variations, *Scientometrics*, 60(3), 475-485.

Tashakkori, A. & Teddlie, C. (1998) Mixed methodology: Combining qualitative and quantitative approaches, Thousand Oaks, CA: Sage Publications Inc.

TEC (2005) Performance based research fund, Available at: http://www.tec.govt.nz/funding/research/pbrf/pbrf.htm (18 Apr. 2005)

The Statistical Cybermetrics Research Group (2001) Database 4: UK university web sites July, 2001, Available at: http://cybermetrics.wlv.ac.uk/database/index.html (12 Nov. 2001)

The Statistical Cybermetrics Research Group (2003) Database 13: UK university web sites June 2003, Available at: http://cybermetrics.wlv.ac.uk/database/index.html (4 Sept. 2003)

The Statistical Cybermetrics Research Group (2005) SocSciBot 3: Link crawler for the social sciences, Available at: http://socscibot.wlv.ac.uk/ (7 July 2005)

Thelwall, M. (2000) Web impact factors and search engine coverage, *Journal of Documentation*, 56(2), 185-189.

Thelwall, M. (2001a) Exploring the link structure of the Web with network diagrams, *Journal of Information Science*, 27(6), 393-401.

Thelwall, M. (2001b) Extracting macroscopic information from web links, *Journal of American Society for Information Science and Technology*, 52(13), 1157-1168.

Thelwall, M. (2001c) A publicly accessible database of UK university website links and a discussion of the need for human intervention in web crawling, University of Wolverhampton. Available at: http://www.eicstes.org/EICSTES_PDF/PAPERS/A%20publicly%20accesible%20database%20of%20UK%20university%20website%20links%20(Thelwall).pdf (9 Oct. 2005)

Thelwall, M. (2001d) The responsiveness of search engine indexes, *Cybermetrics*, 5(1). Available at: http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html.

Thelwall, M. (2001e) Results from a web impact factor crawler, *Journal of Documentation*, 57(2), 177-191.

Thelwall, M. (2001f) A web crawler design for data mining, *Journal of Information Science*, 27(5), 319-325.

Thelwall, M. (2002a) A comparison of sources of links for academic web impact factor calculations, *Journal of Documentation*, 58(1), 66-78.

Thelwall, M. (2002b) Conceptualizing documentation on the Web: An evaluation of different heuristic-based models for counting links between university web sites, *Journal of American Society for Information Science and Technology*, 53(12), 995-1005.

Thelwall, M. (2002c) Evidence for the existence of geographic trends in university web site interlinking, *Journal of Documentation*, 58(5), 563-574.

Thelwall, M. (2002d) An initial exploration of the link relationship between UK university web sites, *ASLIB Proceedings*, 54(2), 118-126.

Thelwall, M. (2002e) Methodologies for crawler based web surveys, *Internet Research: Electronic Networking and Applications and Policy*, 12(2), 124-138.

Thelwall, M. (2002f) A research and institutional size-based model for national university web site interlinking, *Journal of Documentation*, 58(6), 683-694.

Thelwall, M. (2002g) The top 100 linked pages on UK university web sites: high inlink counts are not usually associated with quality scholarly content, *Journal of Information Science*, 28(6), 483-491.

Thelwall, M. (2003a) Can Google's PageRank be used to find the most important academic web pages?, *Journal of Documentation*, 59(2), 205-217.

Thelwall, M. (2003b) Web use and peer interconnectivity metrics for academic web sites, *Journal of Information Science*, 29(1), 1-10.

Thelwall, M. (2003c) What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation, *Information Research*, 8(3). Available at: http://informationr.net/ir/8-3/paper151.html

Thelwall, M. (2004a) Can the Web give useful information about commercial uses of scientific research?, *Online Information Review*, 28(2), 120-130.

Thelwall, M. (2004b) Link analysis: An information science approach, San Diego: Elsevier Academic Press

Thelwall, M. (2005) Text characteristics of English language university web sites, *Journal of the American Society for Information Science and Technology*, 56(6), 609-619.

Thelwall, M. (2005, to appear) Interpreting social science link analysis research: A theoretical framework, *Journal of the American Society for Information Science.*

Thelwall, M., Binns, R., Harries, G., Page-Kennedy, T., Price, E. & Wilkinson, D. (2002) European union associated university websites, *Scientometrics,* 53(1), 95-111.

Thelwall, M. & Harries, G. (2004) Do the web sites of higher rated scholars have significantly more online impact?, *Journal of American Society for Information Science and Technology,* 55(2), 149-159.

Thelwall, M., Harries, G. & Wilkinson, D. (2003a) Why do web sites from different academic subjects interlink?, *Journal of Information Science,* 29(6), 453-471.

Thelwall, M. & Price, L. (2003) Disciplinary differences in academic web presence: A statistical study of the UK, *Libri,* 53(4), 242-253.

Thelwall, M. & Smith, A. G. (2002) Interlinking between Asia-Pacific university web sites, *Scientometrics,* 55(3), 363-376.

Thelwall, M. & Tang, R. (2003) Disciplinary and linguistic considerations for academic web linking: An exploratory hyperlink mediated study with Mainland China and Taiwan, *Scientometrics,* 58(1), 155-181.

Thelwall, M., Tang, R. & Price, E. (2003b) Linguistic patterns of academic web use in Western Europe, *Scientometrics,* 56(3), 417-432.

Thelwall, M., Vaughan, L., Cothey, V., Li, X. & Smith, A. G. (2003c) Which academic subjects have most online impact? A pilot study and a new classification process, *Online Information Review,* 27(5), 333-343.

Thelwall, M. & Wilkinson, D. (2003a) Graph structure in three national academic webs: Power laws with anomalies, *Journal of American Society for Information Science and Technology,* 54(8), 706-712.

Thelwall, M. & Wilkinson, D. (2003b) Three target document range metrics for university web sites, *Journal of American Society for Information Science and Technology,* 54(6), 490-497.

Thelwall, M. & Wilkinson, D. (2004) Finding similar academic web sites with links, bibliometric couplings and colinks, *Information Processing and Management,* 40(3), 515-526.

THES (2005) University performance, *The Higher Education Supplement.* Available at: http://www.thes.co.uk/statistics/university_performance/ (14 Apr. 2005)

Thomas, O. & Willett, P. (2000) Webometric analysis of departments of librarianship and information science, *Journal of Information Science,* 26(6), 421-428.

Tijssen, R. J. W. (2001) Global and domestic utilization of industrial relevant science: Patent citation analysis of science-technology interactions and knowledge flows, *Research Policy,* 30(1), 35-54.

Tijssen, R. J. W., Buter, R. K. & van Leeuwen, T. N. (2000) Technological relevance of science: An assessment of citation linkages between patents and research papers, *Scientometrics,* 47(2), 389-412.

Tognolini, J. A. & Hattie, J. (1994) A methodology to choose performance indicators of research attainment in universities, *Australian Journal of Education,* 38(2), 105-117.

Toutkoushian, R. K., Porter, S. R., Danielson, C. & Hollis, P. R. (2003) Using publications counts to measure an institution's research productivity, *Research in Higher Education,* 44(2), 121-148.

van Raan, A. F. J. (1996) Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises, *Scientometrics,* 36(3), 397-420.

van Raan, A. F. J. (1998) In matters of quantitative studies of science the fault of the theorists is offering too little and asking too much - Comments on theories of citation?, *Scientometrics,* 43(1), 129-139.

van Raan, A. F. J. (2000) The Pandora's box of citation analysis: Measuring scientific excellence-The last evil?, In: Cronin, B. and Atkins, H. B. (eds), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield* Medford, NJ: Information Today, Inc. ASIS Monograph Series, pp. 301-319.

van Raan, A. F. J. (2005) Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods, *Scientometrics,* 62(1), 133-143.

van Raan, A. F. J. & van Leeuwen, T. (2002) Assessment of the scientific basis of interdisciplinary, applied research - Application of bibliometric methods in nutrition and food research, *Research Policy,* 31(4), 611-632.

van Vianen, B. G., Moed, H. & van Raan, A. F. J. (1990) An exploration of the science base of recent technology, *Research Policy,* 19(1), 61-81.

Vaughan, L. (2001) Statistical methods for the information professional, Medford, New Jersey: Information Today.

Vaughan, L. (2002) Web link counts correlate with ISI impact factors: Evidence from two disciplines, In *Proceeding of the 65th ASIST,* Information Today Inc, Medford, Vol.39, pp 436-443.

Vaughan, L. & Thelwall, M. (2003) Scholarly use of the Web: What are the key inducers of links to journal web sites?, *Journal of American Society for Information Science and Technology,* 54(1), 29-38.

Vaughan, L. & Thelwall, M. (2005) A modeling approach to uncover hyperlink patterns: The case of Canadian universities, *Information Processing and Management,* 41(2), 347-359.

Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E. & Deleus, F. (2002) Linking science to technology: Using bibliographic references in patents to build linkage schemes, *Scientometrics,* 54(3), 399-420.

Vinkler, P. (1994) The origin and features of information referenced in pharmaceutical patents, *Scientometrics,* 30(1), 283-302.

Vinkler, P. (1996) Relationships between the rate of scientific development and citations. The chance for citedness model, *Scientometrics,* 35(3), 375-386.

White, H. D. & McCain, K. W. (1998) Visualizing a discipline: An author co-citation analysis of information science, 1972-1995, *Journal of the American Society for Information Science,* 49(4), 327-355.

White, M. D. & Wang, P. L. (1997) A qualitative study of citing behavior: Contributions, criteria, and metalevel documentation concerns, *Library Quarterly,* 67(2), 122-154.

Wilkinson, D., Harries, G., Thelwall, M. & Price, E. (2003) Motivations for academic web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication, *Journal of Information Science,* 29(1), 49-56.

Winship, I. R. (1995) World Wide Web searching tools - An evaluation, *VINE,* 99, 49-54.

Wood, F. Q. (1990) Factors influencing research performance of university academic staff, *Higher Education,* 19(1), 81-100.

Wouters, P. (1999) The citation culture, Amsterdam: University of Amsterdam.

Yitzhaki, M. (1998) The 'language preference' in sociology: Measures of 'language self-citation,' 'relative own language preference indicator,' and 'mutual use of languages', *Scientometrics,* 41(1-2), 243-254.

Zipf, G. K. (1949) Human behavior and the principle of least effort: An introduction to human ecology, Cambridge, MA: Addison-Wesley

Glossary

- All links made: both internal and external outlinks created by a department.
- Alternative Document Models (ADMs). A method of aggregating web content into units for counting purposes. See file ADM, directory ADM, domain ADM and site or department ADM definitions.
- Bibliometrics. The study of the quantitative aspects of the production, dissemination and use of recorded information (Tague-Sutcliffe, 1992).
- Citation. A reference by one publication of another. A citation is the reference viewed from the perspective of the referenced document.
- Co-linked. Web documents or units are linked by same ones. This is an analogy of co-citation in bibliometrics.
- Co-linking, coupling. Web documents or units link to same target . This is an analogy of bibliographic coupling.
- Cybermetrics. The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the whole Internet drawing on bibiometric and informetric approaches (Björneborn, 2004).
- Department ADM. All files belonging to a department are regarded as a single document unit.
- Directory ADM. All files in the same directory are treated as a single document unit. Directories are equated with the position of slashes in URLs, rather than by the actual directory/folder structure of pages on the hosting web server.
- Domain name. The part of an URL of a web page normally following the http:// and preceding the first subsequent slash (if any). Note that this is a simplified definition and there is a longer computer science definition that encompasses additional variations (Berners-Lee et al., 1998).
- Domain ADM. All files with the same domain name are treated as a single document unit.
- File ADM. Each separate file is treated as a document unit for extracting links, or for other counting purposes.
- Host. Used to refer to an individual computer such as a web server.
- Hyperlink. A feature in a web page that allows users to click to navigate to a different web page.
- Informetrics. The study of the quantitative aspects of information in any form, not just records or bibliographics, and in any social group, not just scientists (Tague-Sutcliffe, 1992).
- Inlink. A link to a web unit, a web page or web site (e.g. a web site for a university or department). Unless otherwise stated, inlink means external inlink. External inlink is a link received by a web unit in question from outside. With regard to a web site of a department, an external inlink is defined as a link received from outside the university where the department is hosted.
- Interlink. A link between two different web units. For example, departmental interlinking means links between departments.
- Internet. A large public network of computers running IP and able to communicate with each other.

- Journal Impact Factor (JIF). The number of citations received in the current year by all papers published in a journal during the previous two years, divided by the number of papers published within that period. JIF can also be calculated based on the articles published in the previous one year if the field is a rapidly changing one, or based on the articles of more than two years if the field in question is a less current impact one.
- Link, hyperlink. Links connect web documents together, sometimes referred as 'sitation'. Unless otherwise stated, link can be inlink or outlink regardless of direction.
- Link page. A web page containing a link. This terminology is sometimes used instead of link because search engines count link pages rather than links in response to a link-based query.
- Outlink: a link from a web unit, a web page or web site. Unless otherwise stated, outlink means external outlink. For example, an outlink from a department's web site is a link from the department to a web page outside that university where the department is hosted.
- PageRank. An algorithm used by Google to rank web pages using the link structure of the Web.
- Peers: colleague-scientists in the field.
- Power law. A mathematical law that has been applied to many kinds of web data. It is related to rich-get-richer phenomena. For example, the distribution of TLDs (top level domains) on a given topic follow Lotka's law (Rousseau, 1997).
- Search engine. A program that allows users to type in an information request, such as a keyword query, and returns lists of web pages matching the query.
- Site ADM. All files belonging to a clearly defined web site are treated as a single document unit.
- SocSciBot. A web crawler available online and designed for research crawling.
- SocSciBot Tools. A suite of programs that can be used to analyse the link structure files produced by SocSciBot and those available in cybermetrics university link structure database.
- Top Level Domain (TLD). The final segment of a domain name. This will either be a generic top level domain, such as edu, com and int, or a country-specific domain, such as uk for the UK or ca for Canada.
- Web. The collection of resources that can be obtained over the public Internet using HTTP.
- Web crawler, robot, bot. A program that visits web pages, automatically extracts their links and follows them.
- Web site. A group of related web documents. For example, a personal, departmental or university web site.
- Webometrics. The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and infrometric approaches (Björneborn, 2004).

Appendix A Data for the 79 UK Computer Science Departments

Appendix A1 Non-web Data and Crawler Inlinks

University Name	University Domain	Department Domain	Staff Member	RAE Score	Crawler inlinks
Imperial College of Science, Technology and Medicine	ic.ac.uk	doc.ic.ac.uk	173.4	7	1142
University of Cambridge	cam.ac.uk cambridge.ac.uk	cl.cam.ac.uk	59.3	7	998
University of Edinburgh	ed.ac.uk Edinburgh.ac.uk	dcs.ed.ac.uk dcs.edinburgh.ac.uk informatics.ed.ac.uk cogsci.ed.ac.uk cogsci.edinbugh.ac.uk hcrc.ed.ac.uk aiai.ed.ac.uk dai.ed.ac.uk	232.1	7	3847
University of Manchester	man.ac.uk manchester.ac.uk mcc.ac.uk	cs.man.ac.uk mcc.ac.uk	216.4	7	1110
University of Southampton	soton.ac.uk southampton.ac.uk	ecs.soton.ac.uk	191.4	7	1929
University of York	york.ac.uk	cs.york.ac.uk	154.1	7	1732
Cardiff University	cf.ac.uk cardiff.ac.uk	cs.cardiff.ac.uk cs.cf.ac.uk	109.2	6	350
Lancaster University	lancs.ac.uk Lancaster.ac.uk	comp.lancs.ac.uk	48.9	6	2071
Royal Holloway, University of London	rhul.ac.uk rhbnc.ac.uk	cs.rhul.ac.uk cs.rhbnc.ac.uk dcs.rhbnc.ac.uk	19	6	170
University College London	ucl.ac.uk	cs.ucl.ac.uk	86	6	1478

University of Birmingham	bham.ac.uk birmingham.ac.uk	cs.bham.ac.uk	83.8	6	5082
University of Bristol	bris.ac.uk Bristol.ac.uk	cs.bris.ac.uk compsci.bristol.ac.uk	104.7	6	1125
University of Glasgow	gla.ac.uk glasgow.ac.uk	dcs.gla.ac.uk dcs.glasgow.ac.uk	156.1	6	2719
University of Leeds	leeds.ac.uk	comp.leeds.ac.uk scs.leeds.ac.uk	107.8	6	778
University of Liverpool	liv.ac.uk Liverpool.ac.uk	csc.liv.ac.uk robots.eeng.liv.ac.uk	85.6	6	726
University of Newcastle	ncl.ac.uk newcastle.ac.uk	cs.ncl.ac.uk cs.newcastle.ac.uk csr.nc.ac.uk csr.newcastle.ac.uk dcs.www.ncl.ac.uk dcs.www.newcastle.ac.uk	79.5	6	1015
University of Nottingham	nott.ac.uk nottingham.ac.uk	cs.nott.ac.uk	77.8	6	839
University of Oxford	ox.ac.uk oxford.ac.uk	web.comlab.ox.ac.uk comlab.oxford.ac.uk comlab.ox.ac.uk	53.9	6	3126
University of Plymouth	plym.ac.uk Plymouth.ac.uk	tech.plym.ac.uk/soc tech.plymouth.ac.uk/soc	111	6	128
University of Sheffield	shef.ac.uk Sheffield.ac.uk	dcs.shef.ac.uk	158.2	6	961
University of St Andrews	st-and.ac.uk st-andrews.ac.uk	dcs.st-and.ac.uk dcs.st-andrews.ac.uk	19	6	1264
University of Sussex	susx.ac.uk sussex.ac.uk	cogs.sussex.ac.uk cogs.susx.ac.uk	152.41	6	1559
University of Wales, Swansea	swan.ac.uk Swansea.ac.uk	swan.ac.uk/compsci/ swansea.ac.uk/compsci/	39.7	6	169
University of Warwick	warwick.ac.uk	dcs.warwick.ac.uk	36.8	6	338
Birkbeck College	bbk.ac.uk	dcs.bbk.ac.uk	16.3	5	

City University	city.ac.uk	soi.city.ac.uk cs.city.ac.uk csr.city.ac.uk	96	5	3107
De Montfort University	dmu.ac.uk	cse.dmu.ac.uk cms.dmu.ac.uk	325.9	5	1206
Heriot-Watt University	hw.ac.uk heriot-watt.ac.uk	icbl.hw.ac.uk cee.hw.ac.uk	93.4	5	4149
King's College London	kcl.ac.uk	dcs.kcl.ac.uk	56	5	447
Queen Mary, University of London	qmw.ac.uk qmul.ac.uk	dcs.qmw.ac.uk dcs.qmul.ac.uk tcz.net	71	5	3512
South Bank University	sbu.ac.uk	scism.sbu.ac.uk sbu.ac.uk/scism	113	5	109
The Queen's University of Belfast	qub.ac.uk	cs.qub.ac.uk	117.2	5	490
University of Aberdeen	abdn.ac.uk aberdeen.ac.uk	csd.abdn.ac.uk	21.7	5	299
University of Bath	bath.ac.uk	maths.bath.ac.uk cs.bath.ac.uk	90.1	5	150
University of Bradford	brad.ac.uk bradford.ac.uk	scm.brad.ac.uk scm.bradford.ac.uk comp.brad.ac.uk comp.bradford.ac.uk	77.6	5	314
University of Brighton	bton.ac.uk brighton.ac.uk	snowwhite.it.brighton.ac.uk snowwhite.it.bton.ac.uk it.brighton.ac.uk it.bton.ac.uk itri.bton.ac.uk itri.brighton.ac.uk	81.3	5	383
University of Dundee	dundee.ac.uk	computing.dundee.ac.uk	46.8	5	76
University of Durham	dur.ac.uk durham.ac.uk	dur.ac.uk/computer.science dur.ac.uk/~dcs0www dur.ac.uk/CSM	24.9	5	156

University of East Anglia	uea.ac.uk	sys.uea.ac.uk	43.6	5	227
University of Essex	essex.ac.uk sx.ac.uk	cswww.essex.ac.uk	78.2	5	154
University of Exeter	ex.ac.uk exeter.ac.uk	dcs.ex.ac.uk dcs.exeter.ac.uk secs.ex.ac.uk secs.exeter.ac.uk	59.9	5	322
University of Glamorgan	glam.ac.uk glamorgan.ac.uk	web.glam.ac.uk/schools/soc comp.glam.ac.uk	75.4	5	357
University of Greenwich	gre.ac.uk greenwich.ac.uk	cms1.gre.ac.uk	113	5	6
University of Hertfordshire	herts.ac.uk hertfordshire.ac.uk	feis.herts.ac.uk	216.6	5	275
University of Kent at Canterbury	ukc.ac.uk	cs.ukc.ac.uk	98.1	5	662
University of Leicester	le.ac.uk leicester.ac.uk	mcs.le.ac.uk	31.9	5	478
University of Manchester Institute of Science & Technology	umist.ac.uk	co.umist.ac.uk	216.4	5	41
University of Reading	rdg.ac.uk reading.ac.uk	cs.rdg.ac.uk cs.reading.ac.uk cs-cyb-ee.rdg.ac.uk cs-cyb-ee.reading.ac.uk cyber.rdg.ac.uk cyber.reading.ac.uk elec.rdg.ac.uk elec.reading.ac.uk	52.5	5	1730
University of Ulster	ulst.ac.uk ulster.ac.uk	ulst.ac.uk/cticomp ulster.ac.uk/cticomp ics.ltsn.ac.uk infc.ulst.ac.uk/informatics/cms informatics.ulst.ac.uk informatics.ulster.ac.uk	169.7	5	1223

University of Wales, Aberystwyth	aber.ac.uk aberystwyth.ac.uk	aber.ac.uk/~dcswww/ aber.ac.uk/compsci/	38.8	5	91
Liverpool John Moores University	livjm.ac.uk	cms.livjm.ac.uk	115.8	4	19
Loughborough University	lboro.ac.uk loughborough.ac.uk	lboro.ac.uk/departments/co/ loughborough.ac.uk/departments/co/	116.8	4	6
Manchester Metropolitan University	mmu.ac.uk	doc.mmu.ac.uk docm.mmu.ac.uk	125.7	4	1461
Middlesex University	mdx.ac.uk	cs.mdx.ac.uk	102.4	4	485
Napier University	napier.ac.uk	soc.napier.ac.uk dcs.napier.ac.uk	67	4	2742
Nottingham Trent University	ntu.ac.uk	doc.ntu.ac.uk dcm.ntu.ac.uk	112.8	4	42
Open University	open.ac.uk	mcs.open.ac.uk computing.open.ac.uk	308.3	4	
Robert Gordon University	rgu.ac.uk	scms.rgu.ac.uk	57.4	4	540
University of Huddersfield	hud.ac.uk	helios.hud.ac.uk scom.hud.ac.uk	105.2	4	
University of Hull	hull.ac.uk	dcs.hull.ac.uk	42.6	4	147
University of Paisley	paisley.ac.uk	cis.paisley.ac.uk ces.paisley.ac.uk	150	4	6
University of Strathclyde	strath.ac.uk strathclyde.ac.uk	cs.strath.ac.uk dis.strath.ac.uk cis.strath.ac.uk strath.ac.uk/compsci strath.ac.uk/dis	133.1	4	939
University of Sunderland	sund.ac.uk sunderland.ac.uk	osiris.sund.ac.uk osiris.sunderland.ac.uk cet.sund.ac.uk cet.sunderland.ac.uk hapy.sund.ac.uk hapy.sunderland.ac.uk isis.sund.ac.uk isis.sunderland.ac.uk	99.2	4	268

University of West of England, Bristol	uwe.ac.uk	csm.uwe.ac.uk cems.uwe.ac.uk	136.2	4	
University of Westminster	wmin.ac.uk westminster.ac.uk	cscs.wmin.ac.uk hscs.wmin.ac.uk cpc.wmin.ac.uk	273.7	4	66
Glasgow Caledonian University	gcal.ac.uk	com.gcal.ac.uk	30	3	2
Goldsmiths College	goldsmiths.ac.uk gold.ac.uk	mcs.gold.ac.uk igor.gold.ac.uk	20.4	3	
Keele University	keele.ac.uk	keele.ac.uk/depts/cs	14.8	3	449
Kingston University	king.ac.uk kingston.ac.uk	techweb.king.ac.uk/cses techweb.kingston.ac.uk/cses infosys.king.ac.uk infosys.kingston.ac.uk techweb.king.ac.uk/cis techweb.kingston.ac.uk/cis	71.5	3	
London Guildhall University	lgu.ac.uk	lgu.ac.uk/cism	37.1	3	
Oxford Brookes University	brookes.ac.uk oxford-brookes.ac.uk	wwwcms.brookes.ac.uk wwwcms.oxford-brookes.ac.uk	49.1	3	
Sheffield Hallam University	shu.ac.uk	shu.ac.uk/schools/cms careermsc.com yourmsc.com mscstats.com	306.9	3	90
University of North London	unl.ac.uk	unl.ac.uk/simt	52.9	3	57
University of Stirling	stir.ac.uk stirling.ac.uk	cs.stir.ac.uk	19.1	3	281
Bournemouth University	bournemouth.ac.uk	dec.bournemouth.ac.uk	163.9	2	90
Coventry University	cov.ac.uk coventry.ac.uk	mis.cov.ac.uk mis.coventry.ac.uk	77.9	2	130
University of Northumbria at Newcastle	unn.ac.uk northumbria.ac.uk	computing.unn.ac.uk	144.3	2	
University of Teesside	tees.ac.uk teeside.ac.uk	scm.tees.ac.uk	107.5	2	120
University of Wolverhampton	wlv.ac.uk wolverhampton.ac.uk	scit.wlv.ac.uk	77.5	2	2800

Appendix A2 Number of Inlinks Counted from AltaVista

University Name	External	ac.uk	co.uk	uk	edu	org	com	mil	gov	int	net
Imperial College of Science, Technology and Medicine	62921	16514	2737	19699	3223	8973	10860	6	79	1	5329
University of Cambridge	32470	1632	1018	3103	3166	3915	5013	3	73	1	1800
University of Edinburgh	37120	3208	1459	5050	2620	4835	5909	6	65	1	2085
University of Manchester	18125	1522	1054	3377	868	3054	2806	6	22	0	1208
University of Southampton	25375	2109	644	2873	1727	3235	3593	4	39	0	1665
University of York	7265	1132	216	1442	405	728	1074	3	3	0	319
Cardiff University	4935	426	191	659	474	398	947	1	5	0	236
Lancaster University	13324	1061	297	1730	1249	1279	2052	2	6	1	727
Royal Holloway, University of London	1135	71	17	91	78	119	278	0	2	0	62
University College London	21162	2138	627	3092	1351	2462	2478	3	88	3	1278
University of Birmingham	8349	1005	256	1323	785	821	1328	6	29	0	269
University of Bristol	3326	375	94	484	205	652	623	1	0	0	83
University of Glasgow	683	110	6	127	53	134	98	1	3	0	16
University of Leeds	4301	571	153	1168	559	316	725	0	7	0	135
University of Liverpool	6040	790	163	1022	270	224	807	3	4	0	197
University of Newcastle	3155	234	505	1424	102	179	611	0	1	1	281
University of Nottingham	8241	947	289	1285	512	520	1185	4	4	0	396
University of Oxford	23826	1828	530	2555	2742	2944	5048	3	55	0	1107
University of Plymouth	321	84	10	96	22	20	38	1	3	0	7
University of Sheffield	3717	446	143	627	271	277	494	1	3	1	112

University of St Andrews	4420	1490	61	1567	468	157	581	1		0	402
University of Sussex	10544	1131	263	1440	1738	2204	997	3	58	0	277
University of Wales, Swansea	330	91	8	102	19	17	36	0	0	0	7
University of Warwick	3488	472	129	698	299	286	567	2	2	0	237
Birkbeck College	2408	176	52	236	71	68	188	1	4	0	1450
City University	3053	737	215	1027	605	535	1180	1	6	0	348
De Montfort University	4124	701	217	987	502	290	853	1	6	0	187
Heriot-Watt University	9650	2322	528	3075	697	676	1582	3	31	2	492
King's College London	1510	229	136	375	98	98	206	0	4	1	50
Queen Mary, University of London	5984	549	165	739	357	385	1013	2	3	0	251
South Bank University	1613	185	35	227	117	285	236	1	1	0	68
The Queen's University of Belfast	2355	154	60	226	183	165	435	1	6	0	139
University of Aberdeen	3937	241	62	320	162	148	296	2	3	0	108
University of Bath	1194	244	28	277	154	103	123	0	2	0	26
University of Bradford	887	112	39	165	56	235	162	0	2	0	29
University of Brighton	3129	387	116	524	416	254	533	0	3	0	157
University of Dundee	717	96	65	178	20	26	88	0	0	0	9
University of Durham	263	105	13	121	24	13	27	0	1	0	0
University of East Anglia	4105	524	404	1042	253	317	892	0	8	0	227
University of Essex	852	174	29	210	70	28	83	1	1	0	12
University of Exeter	3916	250	124	474	218	470	1768	1	2	0	100
University of Glamorgan	1145	131	146	294	85	92	254	0	0	0	66
University of Greenwich	162	34	8	46	8	8	40	0	0	0	5
University of Hertfordshire	241	52	18	80	19	18	38	0	0	0	4

University of Kent at Canterbury	2282	392	46	463	542	153	203	2	2	0	47
University of Leicester	815	267	14	284	50	29	106	1	0	0	14
University of Manchester Institute of Science & Technology	1769	238	76	352	95	121	364	0	2	0	95
University of Reading	3401	665	126	856	276	250	529	1	16	0	105
University of Ulster	1675	493	64	590	132	93	237	1	1	0	40
University of Wales, Aberystwyth	328	74	51	138	20	14	72	1	0	0	7
Liverpool John Moores University	567	119	42	171	64	23	133	0	2	0	38
Loughborough University	123	44	7	53	8	12	4	0	0	0	2
Manchester Metropolitan University	3336	491	215	815	307	238	693	2	24	0	122
Middlesex University	602	162	32	211	31	21	126	0	0	0	20
Napier University	10496	487	151	771	282	182	569	1	3	0	83
Nottingham Trent University	399	61	30	95	40	20	81	0	0	0	16
Open University	855	140	95	267	45	47	189	1	1	1	70
Robert Gordon University	1276	128	54	196	46	112	347	2	0	0	98
University of Huddersfield	244	42	18	63	3	2	1	0	0	0	1
University of Hull	3301	156	572	757	150	296	869	1	1	0	340
University of Paisley	251	45	13	64	9	14	42	1	1	0	6
University of Strathclyde	4764	839	258	1215	606	242	242	1	3	3	166
University of Sunderland	3342	304	279	645	296	200	730	2	4	0	215
University of West of England, Bristol	725	112	64	188	55	49	101	1	2	0	64

University of Westminster	614	38	13	52	28	178	123	0	3	1	25
Glasgow Caledonian University	15	4	0	6	0	1	4	0	0	0	0
Goldsmiths College	61	21	1	23	4	3	17	0	0	0	0
Keele University	436	107	22	137	63	19	48	0	2	0	15
Kingston University	75	22	7	33	5	9	5	0	0	0	3
London Guildhall University	44	11	5	18	0	2	9	0	0	0	2
Oxford Brookes University	35	12	2	14	0	2	4	0	0	0	0
Sheffield Hallam University	188	35	13	50	22	8	48	0	0	0	7
University of North London	48	23	2	26	3	2	4	0	0	0	0
University of Stirling	2169	289	129	446	156	75	452	0	2	0	84
Bournemouth University	635	131	34	174	30	43	136	0	0	0	30
Coventry University	465	94	39	149	36	18	68	0	0	0	20
University of Northumbria at Newcastle	86	22	1	25	3	3	16	0	1	0	0
University of Teesside	400	63	41	112	36	12	87	0	1	0	21
University of Wolverhampton	5787	1703	455	2410	254	262	1048	0	1	0	354

Appendix B Data for Australian Physics, Chemistry and Biology Departments

Appendix B1 Non-web Data for Australian Physics Departments

University Name	University domain	Department domain	Staff members	Citation counts	Publications
The Australia National University	anu.edu.au	msowww.anu.edu.au mso.anu.edu.au anu.edu.au/Physics wwwmaths.anu.edu.au/research.programs/aap anu.edu.au/EMU wwwrsphysse.anu.edu.au	115	12090	1340
Curtin University of Technology	curtin.edu.au	physics.curtin.edu.au curtin.edu.au/curtin/dept/phys-sci/apphys/	8	378	90
Flinders University	flinders.edu.au	scieng.flinders.edu.au/cpes ph.flinders.edu.au	29	428	89
James Cook University	jcu.edu.au	jcu.edu.au/school/mathphys/physics	6	36	7
La Trobe University	latrobe.edu.au	latrobe.edu.au/physics	7	174	103
Macquarie University	mq.edu.au	physics.mq.edu.au ics.mq.edu.au/cla	15	864	246
Monash University	monash.edu.au	spme.monash.edu.au physics.monash.edu.au	27	265	121
Murdoch University	murdoch.edu.au	wwwphys.murdoch.edu.au/camsp	10	108	27
Queensland University of Technology	qut.edu.au	pcs.sci.qut.edu.au pcs.sci.qut.com sci.qut.edu.au/aemf	35	464	140
Swinburne University of Technology	swin.edu.au	swin.edu.au/bioscieleceng	51	378	124
University of Adelaide	adelaide.edu.au	physics.adelaide.edu.au	22	2906	500
University of Melbourne	unimelb.edu.au	ph.unimelb.edu.au	41	2670	590
University of New South Wales	unsw.edu.au	phys.unsw.edu.au	30	5567	787
Australian Defence Force Academy	unsw.adfa.edu.au	unsw.adfa.edu.au/physics	12	714	150
University of Newcastle	newcastle.edu.au	newcastle.edu.au/school/math-physical-sci	30	247	88

University of Queensland	uq.edu.au	cmr.uq.edu.au physics.uq.edu.au uq.edu.au/nanoworld	32	3361	610
University of Sydney	usyd.edu.au	physics.usyd.edu.au emu.usyd.edu.au	43	5142	1009
University of Tasmania	utas.edu.au	phys.utas.edu.au/physics phys.utas.edu.au scieng.utas.edu.au/physics	4	659	74
University of Technology, Sydney	uts.edu.au	science.uts.edu.au/physics phys.uts.edu.au/physics	12	138	62
University of Western Sydney	uws.edu.au	uws.edu.au/about/acadorg/cste/seid fistserv.macarthur.uws.edu.au/physics	41	243	21
University of Western Australia	uwa.edu.au	pd.uwa.edu.au physics.uwa.edu.au	12	842	246
University of Wollongong	uow.edu.au	uow.edu.au/eng/phys	9	437	68

Appendix B2 Links at Different Document Levels from SocSciBot for Australian Physics Departments

University Name	F-inlink	Dir-inlink	Dom-inlink	Dept-inlink	F-outlink	Dir-outlink	Dom-outlink	Dept-outlink	Web pages
The Australia National University	147	129	27	8	179	158	58	14	46517
Curtin University of Technology	4	4	4	3	2	1	1	1	374
Flinders University	28	20	11	5	0	0	0	0	281
James Cook University	0	0	0	0	0	0	0	0	103
La Trobe University	0	0	0	0	0	0	0	0	261
Macquarie University	65	47	15	6	50	47	16	8	5063
Monash University	17	17	10	7	1	1	1	1	1015
Murdoch University	3	2	1	1	7	6	3	1	52
Queensland University of Technology	5	3	3	3	4	4	4	3	277
Swinburne University of Technology	2	2	2	2	0	0	0	0	240
University of Adelaide	82	70	24	10	71	60	17	9	4125
University of Melbourne	115	88	28	10	156	136	46	13	33802
University of New South Wales	153	115	29	8	66	64	20	13	10759
Australian Defence Force Academy	1	1	1	1	0	0	0	0	43
University of Newcastle	0	0	0	0	0	0	0	0	120
University of Queensland	38	34	13	6	119	75	20	10	7415
University of Sydney	115	95	21	9	133	90	22	10	7489
University of Tasmania	25	24	17	7	21	20	13	6	3210
University of Technology, Sydney	0	0	0	0	0	0	0	0	19
University of Western Sydney	0	0	0	0	0	0	0	0	28

University of Western Australia	26	25	18	6	24	21	10	7	3322
University of Wollongong	11	9	8	5	4	2	1	1	128

Appendix B3 Inlinks from Different Domains for Australian Physics Departments from AltaVista

University Name	External	com	edu	au	org	net	com.au	edu.au	Web pages
The Australia National University	6278	1180	816	906	263	295	211	456	6873
Central Queensland University	10	1	0	5	0	0	1	3	1
Curtin University of Technology	83	16	12	29	6	2	5	18	1049
Flinders University	244	37	18	90	11	11	4	74	27
James Cook University	18	7	1	0	1	2	0	0	107
La Trobe University	11	1	2	6	0	0	1	1	97
Macquarie University	626	102	108	188	28	13	12	150	1572
Monash University	408	71	38	119	31	19	12	90	395
Murdoch University	34	2	0	22	1	2	1	21	84
Queensland University of Technology	32	6	5	15	0	0	0	12	31
Royal Melborne Institute of Technology	90	5	31	30	1	0	2	24	5
Swinburne University of Technology	113	23	0	45	5	0	9	24	464
University of Adelaide	3469	370	1041	754	320	97	58	626	152
University of Melbourne	4170	1362	282	472	216	668	34	361	11281
University of New South Wales	6103	1639	780	933	745	379	120	667	6821
Australian Defence Force Academy	45	0	0	39	0	0	0	39	8
University of Newcastle	42	3	3	17	6	1	3	8	20
University of Queensland	2834	719	368	312	245	155	45	191	4164
University of Sydney	4908	1305	356	1047	472	587	140	494	4351
University of Tasmania	598	112	39	147	31	21	24	68	3678

University of Technology, Sydney	20	2	1	11	3	0	2	8	23
University of Western Sydney	6	1	0	4	0	0	0	0	28
University of Western Australia	1165	210	79	233	25	17	24	170	1952

Appendix B4 Non-web Data for Australian Chemistry Departments

University Name	University domain	Department domain	Staff members	Citation counts	Publications
Australia National University	anu.edu.au	chemistry.anu.edu.au rsc.anu.edu.au	38	4511	935
Curtin University	curtin.edu.au	chemistry.curtin.edu.au	14	211	73
Flinders University	flinders.edu.au	scieng.flinders.edu.au/cpes	29	316	78
James Cook University	jcu.edu.au	jcu.edu.au/fmhms/school/pms/chem. jcu.edu.au/school/pms/chem	7	266	67
La Trobe University	latrobe.edu.au	latrobe.edu.au/chemistry	9	373	128
Macquarie University	mq.edu.au	chem.mq.edu.au	38	370	103
Monash University	monash.edu.au	chem.monash.edu.au	26	3106	646
Murdoch University	murdoch.edu.au	wwwscience.murdoch.edu.au/chemistry	7	207	64
Queensland University of Technology	qut.edu.au qut.com	pcs.sci.qut.edu.au sci.qut.edu.au/sci_schps	32	995	324
Royal Melbourne Institute of Technology	rmit.edu.au	rmit.edu.au/appchem rmit.edu.au/departments/cm	30	442	145
Southern Cross University	scu.edu.au	scu.edu.au/research/cp/	4	9	26
Swinburne University of Technology	swin.edu.au	hed.swin.edu.au/ses	35	74	8
University of Adelaide	adelaide.edu.au	chemistry.adelaide.edu.au	9	1677	512
University of Melbourne	unimelb.edu.au	chemistry.unimelb.edu.au	22	2742	537
The University of New England	une.edu.au	une.edu.au/chemistry une.edu.au/~chemistry	8	247	48
University of New South Wales	unsw.edu.au	chem.unsw.edu.au	19	1840	466
Australian Defence Force Academy	adfa.edu.au	unsw.adfa.edu.au/chemistry ch.adfa.edu.au	8	235	71
University of Newcastle	newcastle.edu.au	newcastle.edu.au/discipline/chemistry	10	279	89
University of Queensland	uq.edu.au	chemistry.uq.edu.au ilc00f.facbacs.uq.edu.au/smms/chem	26	1448	382

University of South Australia	unisa.edu.au	unisa.edu.au/iwri unisa.edu.au/pmbs	62	717	244
University of Sydney	usyd.edu.au	chem.usyd.edu.au	38	3328	677
University of Tasmania	utas.edu.au	scieng.utas.edu.au/chem. chem.utas.edu.au	12	1217	202
University of Western Australia	uwa.edu.au	donotpublish.weboffice.uwa.edu.au/chemistry chem.uwa.edu.au	16	2028	464
University of Western Sydney	uws.edu.au	uws.edu.au/about/acadorg/cste/ssfh	66	159	32
University of Wollongong	uow.edu.au	uow.edu.au/science/chem	18	810	172

Appendix B5 Links at Different Document Levels from SocSciBot for Australian Chemistry Departments

University Name	F-inlink	Dir-inlink	Dom-inlink	Dept-inlink	F-outlink	Dir-outlink	Dom-outlink	Dept-outlink	Web pages
Australia National University	6	6	3	3	51	37	15	12	7606
Curtin University	1	1	1	1	0	0	0	0	349
Flinders University	0	0	0	0	0	0	0	0	281
James Cook University	1	1	1	1	5	5	3	3	163
La Trobe University	0	0	0	0	0	0	0	0	130
Macquarie University	9	7	3	3	5	5	5	4	496
Monash University	5	5	4	4	2	2	2	2	1109
Murdoch University	0	0	0	0	0	0	0	0	3
Queensland University of Technology	0	0	0	0	0	0	0	0	227
Royal Melbourne Institute of Technology	3	3	2	2	0	0	0	0	2
Southern Cross University	0	0	0	0	0	0	0	0	12
Swinburne University of Technology	0	0	0	0	0	0	0	0	172
University of Adelaide	1	1	1	1	0	0	0	0	477
University of Melbourne	8	6	4	4	9	8	3	3	1727
The University of New England	2	2	1	1	0	0	0	0	955
University of New South Wales	8	7	5	5	10	9	6	5	1017
Australian Defence Force Academy	5	4	4	2	0	0	0	0	69
University of Newcastle	4	1	1	1	2	2	2	2	187
University of Queensland	4	3	2	2	0	0	0	0	45
University of South Australia	0	0	0	0	0	0	0	0	220

University of Sydney	29	23	8	5	8	8	8	8	2917
University of Tasmania	4	4	3	3	0	0	0	0	10
University of Western Australia	1	1	1	1	1	1	1	1	424
University of Western Sydney	0	0	0	0	0	0	0	0	166
University of Wollongong	2	2	1	1	0	0	0	0	99

Appendix B6 Inlinks from Different Domains for Australian Chemistry Departments from AltaVista

University Name	External	com	edu	au	org	net	com.au	edu.au	Web pages
Australia National University	1242	96	93	692	34	16	20	624	110
Curtin University	100	13	3	44	1	0	5	29	226
Flinders University	83	13	7	37	5	4	2	32	27
James Cook University	6	2	0	3	0	0	0	2	46
La Trobe University	11	2	1	2	0	0	0	0	107
Macquarie University	270	23	23	114	9	3	7	95	333
Monash University	319	57	36	115	5	5	9	82	238
Murdoch University	34	4	1	23	1	0	0	23	1
Queensland University of Technology	31	1	1	23	0	0	5	16	1
Royal Melbourne Institute of Technology	109	13	6	33	9	1	6	23	88
Southern Cross University	10	4	1	1	1	1	1	0	10
Swinburne University of Technology	89	14	0	67	1	0	4	56	119
University of Adelaide	152	25	12	67	6	2	2	39	311
University of Melbourne	394	93	25	135	6	6	8	114	705
The University of New England	29	4	4	18	3	0	2	14	463
University of New South Wales	826	150	73	347	1	29	17	310	565
Australian Defence Force Academy	358	22	24	91	12	3	1	58	3181
University of Newcastle	30	3	2	8	7	0	0	7	36
University of Queensland	115	4	5	77	3	2	6	45	30
University of South Australia	13	0	2	6	0	0	3	1	96
University of Sydney	447	84	30	169	25	6	11	127	1721

University of Tasmania	2646	44	30	2415	9	14	10	2390	3
University of Western Australia	190	27	7	111	9	1	4	99	478
University of Western Sydney	8	2	0	5	0	0	1	3	172
University of Wollongong	45	3	4	16	3	2	2	12	116

Appendix B7 Non-web Data for Australian Biology Departments

University Name	University domain	Department domain	Staff members	Citation counts	Publications
Australia National University	anu.edu.au	jcsmr.anu.edu.au/dicb jcsmr.anu.edu.au/group_pages/mgc rsbs.anu.edu.au biology.anu.edu.au anu.edu.au/bambi anu.edu.au/BoZo	47	6522	977
University of Central Queensland	cqu.edu.au	ahs.cqu.edu.au/cma	2	249	63
Charles Darwin university	cdu.edu.au	ntu.edu.au/faculties/site/schools/becs	21	24	12
Curtin University	curtin.edu.au	envbio.curtin.edu.au wbiomed.curtin.edu.au/teach/humanbiol	15	199	76
Edith Cowan University	cowan.edu.au	chg.ecu.edu.au	6	412	41
Flinders University	flinders.edu.au	scieng.flinders.edu.au/biology	31	1162	258
James Cook University	jcu.edu.au	*jcu.edu.au/fmhms/school/pms/bchemmbiol* *actfr.jcu.edu.au* jcu.edu.au/school/mbiolaq jcu.edu.au/school/bsc	54	2130	460
La Trobe University	latrobe.edu.au	latrobe.edu.au/bendigo/brc latrobe.edu.au/biochemistry latrobe.edu.au/botany latrobe.edu.au/genetics latrobe.edu.au/microbiology zoo.latrobe.edu.au	53	3208	584
Macquarie University	mq.edu.au	bio.mq.edu.au	35	3051	494
Monash University	monash.edu.au	biolsci.monash.edu.au	18	1237	286
Murdoch University	murdoch.edu.au	murdoch.edu.au/biology murdoch.edu.au/biotech murdoch.edu.au/conbiol murdoch.edu.au/marsci murdoch.edu.au/molbiol	30	1104	227
Queensland University of Technology	qut.edu.au	life.sci.qut.edu.au life.sci.qut.edu.au/html/research/molbiol(old)	36	435	94
Royal Melbourne Institute of Technology	rmit.edu.au	rmit.edu.au/biotechnology	17	183	70

University of Adelaide	adelaide.edu.au	cmgd.adelaide.edu.au adelaide.edu.au/sciences/env_biol/research/fresh ees.adelaide.edu.au/research/enviro mbs.adelaide.edu.au	113	4437	762
University of Canberra	canberra.edu.au	aerg.canberra.edu.au canberra.edu.au/WWW/www-crcfe.nsf	13	340	77
University of Melbourne	unimelb.edu.au	genetics.unimelb.edu.au zoology.unimelb.edu.au botany.unimelb.edu.au cab.unimelb.edu.au	44	4527	797
University of New England	une.edu.au	une.edu.au/botany une.edu.au/bbms	35	1031	207
University of New South Wales	unsw.edu.au	babs.unsw.edu.au	8	1388	233
Newcastle University	newcastle.edu.au	newcastle.edu.au/discipline/biology	16	418	85
University of Queensland	uq.edu.au	tpp.uq.edu.au botany.uq.edu.au zen.uq.edu.au sols.uq.edu.au ccb.uq.edu.au biosci.uq.edu.au/micro/academic/sly/cbdi(old) biosci.uq.edu.au/Html/StaffInterests/CPSFE(old) bacs.uq.edu.au smms.uq.edu.au imb.uq.edu.au	330	10232	1668
University of Sydney	usyd.edu.au	proteomics.usyd.edu.au bio.usyd.edu.au mmb.usyd.edu.au	64	6936	1032
University of Tasmania	utas.edu.au	scieng.utas.edu.au/zoo	10	495	198
University of Technology Sydney	uts.edu.au	science.uts.edu.au/cmb science.uts.edu.au/centres/mau	16	707	160

University of Western Australia	uwa.edu.au	biochem.uwa.edu.au general.uwa.edu.au/u/biochem(old) microbiol.uwa.edu.au animals.uwa.edu.au plants.uwa.edu.au botany.uwa.edu.au	65	4072	790
University of Western Sydney	uws.edu.au	uws.edu.au/about/acadorg/cste/ssfh	66	197	59
University of Wollongong	uow.edu.au	uow.edu.au/science/research/icb/antdiary uow.edu.au/science/biol	17	786	148

Appendix B8 Links at Different Document Levels from SocSciBot for Australian Biology Departments

University Name	F-inlink	Dir-inlink	Dom-inlink	Dept-inlink	F-outlink	Dir-outlink	Dom-outlink	Dept-outlink	Web pages
Australia National University	21	18	12	10	22	20	15	12	1131
University of Central Queensland	0	0	0	0	0	0	0	0	175
Charles Darwin university	0	0	0	0	3	2	2	2	258
Curtin University	0	0	0	0	0	0	0	0	393
Edith Cowan University	0	0	0	0	0	0	0	0	43
Flinders University	1	1	1	1	0	0	0	0	166
James Cook University	1	1	1	1	1	1	1	1	331
La Trobe University	4	4	3	3	9	5	5	5	445
Macquarie University	15	11	6	6	11	11	5	5	1001
Monash University	0	0	0	0	6	6	3	3	238
Murdoch University	0	0	0	0	0	0	0	0	81
Queensland University of Technology	1	1	1	1	2	1	1	1	5855
Royal Melbourne Institute of Technology	0	0	0	0	0	0	0	0	8
University of Adelaide	0	0	0	0	1	1	1	1	193
University of Canberra	7	7	5	4	0	0	0	0	10391
University of Melbourne	7	7	4	4	10	8	8	5	1051
University of New England	0	0	0	0	0	0	0	0	58
University of New South Wales	0	0	0	0	6	6	3	3	276
Newcastle University	2	2	2	2	4	2	1	1	176
University of Queensland	12	8	8	5	4	4	4	3	2984
University of Sydney	11	9	6	5	9	9	5	5	4653
University of Tasmania	1	1	1	1	0	0	0	0	10

University of Technology Sydney	0	0	0	0	0	0	0	0	28
University of Western Australia	2	2	2	2	2	1	1	1	2480
University of Western Sydney	0	0	0	0	0	0	0	0	166
University of Wollongong	5	5	3	3	0	0	0	0	417

Appendix B9 Inlinks from Different Domains for Australian Biology Departments from AltaVista

University Name	External	com	edu	au	org	net	com.au	edu.au	Web pages
Australia National University	2443	230	1281	328	154	11	42	184	743
University of Central Queensland	13	1	0	9	0	0	0	9	211
Charles Darwin university	17	2	0	15	0	0	1	12	162
Curtin University	72	3	12	27	4	0	2	19	320
Edith Cowan University	20	2	1	12	4	0	1	11	52
Flinders University	85	9	7	48	3	2	4	34	212
James Cook University	151	24	25	35	12	10	6	12	282
La Trobe University	323	90	24	62	31	13	10	26	499
Macquarie University	689	75	95	294	30	29	60	183	1055
Monash University	25	0	0	25	0	0	0	25	1
Murdoch University	25	8	2	8	0	1	3	3	33
Queensland University of Technology	625	107	148	93	38	21	9	62	1315
Royal Melbourne Institute of Technology	7	2	1	4	0	0	2	2	10
University of Adelaide	36	12	1	8	2	1	1	2	230
University of Canberra	932	206	107	284	75	32	43	134	391
University of Melbourne	639	102	65	350	25	9	25	265	1044
University of New England	82	6	12	28	6	4	1	15	84
University of New South Wales	389	9	8	352	7	0	4	341	240
Newcastle University	45	11	0	18	7	0	3	12	180
University of Queensland	476	75	94	658	72	23	42	486	1480
University of Sydney	1142	167	124	487	103	25	22	381	2075
University of Tasmania	21	2	0	16	0	0	1	14	4
University of Technology Sydney	57	3	0	46	1	0	4	41	0

University of Western Australia	498	65	22	316	10	15	11	273	2320
University of Western Sydney	8	2	0	5	0	0	1	3	172
University of Wollongong	79	3	10	37	4	6	8	16	623

Appendix C Data for Canadian Physics, Chemistry and Biology Departments

Appendix C1 Non-web Data for Canadian Physics Departments

University Name	University domain	Department domain	Staff members	Research grants	Citation counts	Publications
University of Alberta	ualberta.ca	phys.ualberta.ca	39	946740	7641	515
Brandon University	brandonu.ca	brandonu.ca/physics brandonu.ca/Physics	6	35000	248	51
University of British Columbia	ubc.ca	astro.ubc.ca physics.ubc.ca	71	675308	13560	702
Brock University	brocku.ca	physics.brocku.ca	7	58000	331	67
University of Calgary	ucalgary.ca	phas.ucalgary.ca	24	122300	2156	206
Carleton University	carleton.ca	physics.carleton.ca	13	292000	2865	222
University of Concordia	concordia.ca	physics.concordia.ca	7	45000	797	167
University of Dalhousie	dal.ca	physics.dal.ca	21	275687	1040	130
Université de Moncton	umoncton.ca	umoncton.ca/genie/electrique	7	18000	306	30
Université de Sherbrooke	usherbrooke.ca usherb.ca	usherbrooke.ca/physique physique.usherbrooke.ca	11	173000	2143	205
École Polytechnique	polymtl.ca	phys.polymtl.ca	18	159300	1306	163
Universtiy of Guelph	uoguelph.ca	physics.uoguelph.ca	20	522000	1786	202
Institut national de recherche scientifique	inrs.uquebec.ca	inrs-ener.uquebec.ca	35	741381	941	88
Lakehead University	lakeheadu.ca	lakeheadu.ca/~physwww	4	21300	134	26
Université Laval	ulaval.ca	phy.ulaval.ca	23	264750	2407	215
University of Manitoba	umanitoba.ca	physics.umanitoba.ca	23	253000	1082	133
McGill University	mcgill.ca	physics.mcgill.ca	42	817894	9705	606
Mcmaster University	mcmaster.ca	physics.mcmaster.ca physwww.mcmaster.ca physun.mcmaster.ca	30	965658	5095	435
Memorial University of Newfoundland	mun.ca	physics.mun.ca	19	58800	444	64
University of Montréal	umontreal.ca	phys.umontreal.ca	33	883307	6115	565

Mount Allison University	mta.ca	mta.ca/faculty/science/physics	6	65000	50	12
University of New Brunswick	unb.ca unbsj.ca	unb.ca/fredericton/science/physics	10	43500	610	106
University of Ottawa	uottawa.ca	science.uottawa.ca/phy physics.uottawa.ca	31	236500	1670	189
University of Prince Edward Island	upei.ca	upei.ca/~physics upei.ca/physics	5	15000	72	12
Québec à Trois-Rivières	uqtr.ca	uqtr.ca/Departement/physique uqtr.ca/dphy	7	13000	180	25
Queen's University at Kingston	queensu.ca	physics.queensu.ca phy.queensu.ca	28	299870	3268	352
University of Regina	uregina.ca	phys.uregina.ca	9	110000	1380	119
Ryerson University	ryerson.ca	mpcs.ryerson.ca	46	48000	63	21
University of Saint Mary's	stmarys.ca	ap.stmarys.ca apwww.stmarys.ca	10	114840	578	33
University of Saskatchewan	usask.ca	physics.usask.ca	24	135000	584	97
Simon Fraser University	sfu.ca	sfu.ca/physics sfu.ca/phys phys.sfu.ca	29	815000	3685	353
University of Toronto	utoronto.ca	physics.utoronto.ca astro.utoronto.ca cita.utoronto.ca	98	1787878	6697	493
University of Victoria	uvic.ca	phys.uvic.ca	19	710000	3439	222
University of Waterloo	uwaterloo.ca	science.uwaterloo.ca/physics	32	171800	4092	499
University of Western Ontario	uwo.ca	physics.uwo.ca	31	318000	1654	191
University of Winnipeg	uwinnipeg.ca	physics.uwinnipeg.ca	7	92000	274	21
York University	yorku.ca	physics.yorku.ca science.yorku.ca/units/phas	23	477730	2910	273

Appendix C2 Links at Different Document Levels from SocSciBot for Canadian Physics Departments

University Name	F-inlink	Dir-inlink	Dom-inlink	Dept-inlink	F-outlink	Dir-outlink	Dom-outlink	Dept-outlink	Web pages
University of Alberta	61	51	39	15	47	46	31	17	5730
Brandon University	6	2	1	1	0	0	0	0	167
University of British Columbia	117	99	50	30	187	114	54	17	39877
Brock University	16	15	9	9	51	32	16	11	5184
University of Calgary	19	17	10	9	9	9	8	8	568
Carleton University	52	31	13	11	72	41	21	13	4257
University of Concordia	3	3	3	3	0	0	0	0	27
University of Dalhousie	9	9	7	7	40	32	15	14	432
Université de Moncton	0	0	0	0	0	0	0	0	122
Université de Sherbrooke	5	5	2	2	15	15	11	11	819
École Polytechnique	16	15	7	5	2	2	2	2	525
Universtiy of Guelph	90	74	30	21	35	18	10	9	3710
Institut national de recherche scientifique	3	3	3	3	9	9	6	6	374
Lakehead University	1	1	1	1	5	5	4	4	333
Université Laval	12	11	9	6	11	9	4	4	2334
University of Manitoba	13	13	9	7	19	14	12	11	3247
McGill University	54	50	27	19	0	0	0	0	256
Mcmaster University	65	47	29	16	79	66	48	27	4681
Memorial University of Newfoundland	18	16	9	7	29	25	18	12	3740
University of Montréal	5	4	4	4	10	9	4	3	295
Mount Allison University	23	20	3	3	3	3	3	3	64
University of New Brunswick	0	0	0	0	4	4	4	4	162
University of Ottawa	21	8	8	7	0	0	0	0	105
University of Prince Edward Island	8	8	5	5	1	1	1	1	29

Québec à Trois-Rivières	0	0	0	0	0	0	0	0	20
Queen's University at Kingston	156	97	50	26	69	37	19	14	2546
University of Regina	3	2	2	1	40	40	38	26	1231
Ryerson University	0	0	0	0	3	3	3	3	2391
University of Saint Mary's	30	21	10	10	8	8	5	5	460
University of Saskatchewan	7	7	7	7	36	33	12	12	5416
Simon Fraser University	34	28	16	14	9	8	4	4	3269
University of Toronto	211	160	80	39	185	160	74	38	27003
University of Victoria	70	60	39	17	72	62	30	14	18969
University of Waterloo	21	13	10	8	8	5	1	1	444
University of Western Ontario	8	8	8	7	34	30	18	11	3360
University of Winnipeg	0	0	0	0	14	13	12	11	28
York University	19	16	10	9	70	61	22	13	273

Appendix C3 Inlinks from Different Domains for Canadian Physics Departments from AltaVista

University Name	External	com	edu	ca	org	net	Web pages
University of Alberta	2379	546	354	471	153	125	4497
Brandon University	17	1	1	8	0	0	90
University of British Columbia	4830	1005	711	807	409	241	10795
Brock University	549	140	85	86	39	16	1659
University of Calgary	637	49	31	473	17	12	610
Carleton University	1033	175	156	296	45	41	2273
University of Concordia	65	9	12	31	0	1	84
University of Dalhousie	139	33	12	70	8	0	520
Université de Moncton	9	1	4	3	0	0	17
Université de Sherbrooke	7	1	0	3	0	0	738
École Polytechnique	106	18	11	37	1	2	304
Universtiy of Guelph	2909	687	566	433	228	165	2825
Institut national de recherche scientifique	467	108	62	95	29	15	96
Lakehead University	34	5	6	22	0	0	385
Université Laval	567	120	71	98	27	25	566
University of Manitoba	538	45	96	202	19	11	2551
McGill University	3329	564	649	307	228	172	603
Mcmaster University	1862	306	231	459	160	66	2307
Memorial University of Newfoundland	1134	521	107	159	72	158	2599
University of Montréal	117	10	5	51	9	2	42
Mount Allison University	32	3	10	17	0	0	74
University of New Brunswick	5	1	1	2	1	0	95
University of Ottawa	157	14	11	73	6	7	169
University of Prince Edward Island	145	72	7	31	7	8	699
Québec à Trois-Rivières	1	0	0	1	0	0	4
Queen's University at Kingston	1866	266	345	411	135	57	1946
University of Regina	289	106	33	57	28	11	295

Ryerson University	21	2	1	18	0	0	241
University of Saint Mary's	759	156	68	260	65	38	2456
University of Saskatchewan	413	60	44	204	13	8	1853
Simon Fraser University	600	149	124	70	35	21	1073
University of Toronto	6353	1141	1306	1044	526	173	8650
University of Victoria	2046	323	315	274	443	72	5061
University of Waterloo	472	46	33	275	23	11	356
University of Western Ontario	785	165	109	223	57	28	1891
University of Winnipeg	5	0	0	3	2	0	20
York University	268	29	34	105	12	4	162

Appendix C4 Non-web Data for Canadian Chemistry Departments

University Name	University domain	Department domain	Staff members	Research grants	Citation counts	Publications
Acadia University	acadiau.ca	acadiau.ca/science/chem	12	36300	402	39
University of Alberta	ualberta.ca	chem.ualberta.ca	26	711290	7615	564
University of British Columbia	ubc.ca	chem.ubc.ca	52	1018401	7461	639
University of Brock	brocku.ca	brocku.ca/chemistry	16	174451	727	86
University of Calgary	ucalgary.ca	chem.ucalgary.ca	25	753526	6292	428
Carleton University	carleton.ca	carleton.ca/chem	20	113000	1663	123
Concordia University	concordia.ca	concordia.ca/chem	21	185000	1260	101
Dalhousie University	dal.ca	chem.dal.ca	26	206571	2130	226
Université de Sherbrooke	usherbrooke.ca usherb.ca	usherbrooke.ca/chimie	16	152600	915	117
Québec à Montréal	uqam.ca	er.uqam.ca/nobel/dep_chim	25	212180	629	73
University of Guelph	uoguelph.ca	chembio.uoguelph.ca	38	465613	2443	288
The King's University College	kingsu.ab.ca	kingsu.ab.ca/~chem	4	20000	0	0
Lakehead University	lakeheadu.ca	lakeheadu.ca/~chemdeptwww	7	37300	123	24
Université Laval	ulaval.ca	Chm.ulaval.ca	15	254900	2948	232
University of Lethbridge	uleth.ca/	uleth.ca/chm uleth.ca/fas/chm	11	76300	149	29
University of Manitoba	umanitoba.ca	umanitoba.ca/faculties/science/chemistry umanitoba.ca/chemistry	26	382965	1655	141
McGill University	mcgill.ca	mcgill.ca/chemistry	35	985619	4037	275
McMaster University	mcmaster.ca	chemistry.mcmaster.ca	35	363218	3081	312
University of Montréal	umontreal.ca	chimie.umontreal.ca	31	620331	3918	326
Mount Allison University	mta.ca	Mta.ca/faculty/science/chem	9	20000	40	6
University of New Brunswick	unb.ca unbsj.ca	Unb.ca/fredericton/science/chem	19	50000	668	87
University of Northern British Columbia	unbc.ca	unbc.ca/chemistry	7	20000	20	8
University of Ottawa	uottawa.ca	chem.uottawa.ca	29	426620	4698	427

University of Prince Edward Island	upei.ca	upei.ca/~chem	7	23000	238	34
Queen's University	queensu.ca	chem.queensu.ca	26	456969	2696	298
Québec à Rimouski	uqar.qc.ca uqar.uquebec.ca	uqar.uquebec.ca/bcss uqar.qc.ca/bcss	19	20000	53	5
University of Regina	uregina.ca	chem.uregina.ca	11	267860	301	41
Ryerson University	ryerson.ca	ryerson.ca/cab ryerson.ca/~cab ryerson.ca/~cbce ryerson.ca/chemeng	25	57983	53	6
Saint Mary's University	stmarys.ca	stmarys.ca/academic/science/chemist	8	41300	483	32
University of Saskatchewan	usask.ca	usask.ca/chemistry	18	433470	1256	170
Simon Fraser University	sfu.ca	sfu.ca/chemistry	21	299600	978	149
St. Francis Xavier University	stfx.ca	Stfx.ca/people/chemist	11	32000	361	60
University of Toronto	utoronto.ca	chem.utoronto.ca chem-eng.utoronto.ca	101	1610906	7986	645
Trent University	trentu.ca	trentu.ca/chemistry	7	110300	468	67
University of Victoria	uvic.ca	chemistry.uvic.ca	20	302407	693	98
University of Waterloo	uwaterloo.ca	uwaterloo.ca/chemistry	30	541954	5381	449
University of Western Ontario	uwo.ca	Uwo.ca/chem	37	667581	2406	266
Wilfrid Laurier University	wlu.ca	wlu.ca/~wwwchem	10	43000	274	15
University of Winnipeg	uwinnipeg.ca	chemistry.uwinnipeg.ca	16	21000	651	23
York University	yorku.ca	chem.yorku.ca	30	393954	1525	161

Appendix C5 Links at Different Document Levels from SocSciBot for Canadian Chemistry Departments

University Name	F-inlink	Dir-inlink	Dom-inlink	Dept-inlink	F-outlink	Dir-outlink	Dom-outlink	Dept-outlink	Web pages
Acadia University	5	4	3	3	12	12	11	10	100
University of Alberta	53	25	13	10	33	28	19	13	4869
University of British Columbia	55	38	17	13	3	3	3	3	2815
University of Brock	14	4	3	3	0	0	0	0	68
University of Calgary	30	22	12	10	458	82	32	25	15707
Carleton University	3	3	3	3	0	0	0	0	159
Concordia University	11	5	2	2	1	1	1	1	526
Dalhousie University	51	28	15	10	9	7	7	7	259
Université de Sherbrooke	1	1	1	1	0	0	0	0	111
Québec à Montréal	0	0	0	0	2	2	1	1	82
University of Guelph	428	72	15	11	11	10	7	7	2754
The King's University College	1	1	1	1	0	0	0	0	16
Lakehead University	1	1	1	1	11	9	6	6	62
Université Laval	16	9	7	6	14	14	6	5	3193
University of Lethbridge	3	3	3	3	1	1	1	1	124
University of Manitoba	13	11	7	7	1	1	1	1	722
McGill University	24	11	7	7	0	0	0	0	351
McMaster University	24	24	15	14	331	146	44	39	2826
University of Montréal	8	6	4	4	6	3	3	3	346
Mount Allison University	5	4	4	4	24	20	15	15	120
University of New Brunswick	5	5	4	4	5	4	4	4	300
University of Northern British Columbia	1	1	1	1	0	0	0	0	734
University of Ottawa	8	6	3	2	1	1	1	1	217

University of Prince Edward Island	8	5	4	4	0	0	0	0	136
Queen's University	51	30	17	14	5	5	4	4	1422
Québec à Rimouski	0	0	0	0	0	0	0	0	119
University of Regina	9	6	3	3	0	0	0	0	28
Ryerson University	2	2	2	2	0	0	0	0	138
Saint Mary's University	3	3	3	3	5	5	5	5	32
University of Saskatchewan	14	8	4	4	1	1	1	1	353
Simon Fraser University	18	16	9	9	13	9	9	8	4546
St. Francis Xavier University	8	4	4	4	0	0	0	0	161
University of Toronto	56	23	10	8	31	23	16	12	4463
Trent University	5	5	4	4	4	4	4	4	165
University of Victoria	15	8	3	3	45	35	16	15	1317
University of Waterloo	37	18	11	11	4	3	2	2	435
University of Western Ontario	52	23	11	11	3	3	3	3	609
Wilfrid Laurier University	1	1	1	1	6	5	4	4	227
University of Winnipeg	1	1	1	1	0	0	0	0	157
York University	17	14	9	9	17	14	11	11	1454

Appendix C6 Inlinks from Different Domains for Canadian Chemistry Departments from AltaVista

University Name	External	com	edu	ca	org	net	Web pages
Acadia University	62	6	6	30	2	0	0
University of Alberta	2110	383	391	512	140	85	2435
University of British Columbia	511	103	73	213	17	14	735
University of Brock	56	11	5	26	3	0	240
University of Calgary	1199	222	146	461	50	34	6136
Carleton University	39	6	3	17	1	0	37
Concordia University	58	5	3	26	3	1	161
Dalhousie University	376	43	36	190	13	8	258
Université de Sherbrooke	18	1	0	14	1	0	4
Québec à Montréal	12	3	1	2	1	0	3
University of Guelph	1827	452	356	489	88	73	3390
The King's University College	16	1	2	6	1	0	141
Lakehead University	98	3	2	84	1	0	67
Université Laval	221	80	12	44	6	3	600
University of Lethbridge	58	13	6	13	4	3	105
University of Manitoba	149	17	14	54	5	3	431
McGill University	135	20	22	55	9	3	10
McMaster University	1357	235	286	228	68	40	1786
University of Montréal	71	6	3	46	1	0	216
Mount Allison University	53	8	3	23	0	4	113
University of New Brunswick	9	1	0	6	0	2	195
University of Northern British Columbia	24	5	1	9	1	2	196
University of Ottawa	179	29	24	31	2	2	157
University of Prince Edward Island	82	4	1	62	0	0	30
Queen's University	541	79	90	152	18	13	1792
Québec à Rimouski	7	0	0	6	0	1	13
University of Regina	57	5	4	37	1	0	21
Ryerson University	19	8	0	7	0	0	90

Saint Mary's University	83	10	5	44	1	2	28
University of Saskatchewan	103	7	9	26	7	1	180
Simon Fraser University	199	82	18	47	5	3	617
St. Francis Xavier University	35	5	1	17	0	1	109
University of Toronto	1656	478	243	428	112	42	2825
Trent University	45	6	4	23	2	2	116
University of Victoria	164	16	11	91	9	3	1045
University of Waterloo	167	25	15	66	7	8	328
University of Western Ontario	185	27	24	65	4	5	126
Wilfrid Laurier University	92	3	1	88	0	0	65
University of Winnipeg	5	1	0	4	0	0	0
York University	312	67	23	136	14	14	608

Appendix C7 Non-web Data for Canadian Biology Departments

University Name	University domain	Department domain	Staff members	Research grants	Citation counts	Publications
Acadia University	acadiau.ca	acadiau.ca/science/biol	19	109680	200	53
University of Alberta	ualberta.ca	biology.ualberta.ca afns.ualberta.ca	113	1043231	8426	775
Brandon University	brandonu.ca	brandonu.ca/Zoology brandonu.ca/botany	10	59650	132	21
University of British Columbia	ubc.ca	zoology.ubc.ca microbiology.ubc.ca physiology.ubc.ca biochem.ubc.ca botany.ubc.ca	155	764465	34838	2090
University of Brock	brocku.ca	brocku.ca/biology	19	254145	1237	153
University of Calgary	ucalgary.ca	bio.ucalgary.ca cell.ucalgary.ca ucalgary.ca/UofC/faculties/medicine/PHBI md.ucalgary.ca/Medicine/Departments/MID	133	560310	22832	1676
Concordia University	concordia.ca	concordia.ca/biology	20	92400	3260	255
Dalhousie University	dal.ca	biochem.dal.ca physiology.dal.ca dal.ca/~biology2	85	299497	6030	531
University of Guelph	uoguelph.ca	aps.uoguelph.ca uoguelph.ca/biomed uoguelph.ca/hb+ns uoguelph.ca/zoology uoguelph.ca/microbiology uoguelph.ca/OAC/foodsci uoguelph.ca/mbgwww plant.uoguelph.ca uoguelph.ca/OAC/env uoguelph.ca/botany	228	1579467	15891	1762
Lakehead University	lakeheadu.ca	lakeheadu.ca/~biowww	12	16500	179	42

Laurentian University	laurentian.ca	laurentian.ca/biology	22	48354	845	111
Université Laval	ulaval.ca	ulaval.ca/fmed/bcx sbf.ulaval.ca san.ulaval.ca fmed.ulaval.ca/ap bcm.ulaval.ca	170	620853	15995	1346
University of Lethbridge	uleth.ca	uleth.ca/bio	15	223100	2276	151
University of Manitoba	Umanitoba.ca	umanitoba.ca/faculties/human_ecology/foods umanitoba.ca/human_ecology/foods umanitoba.ca/faculties/science/zoology umanitoba.ca/science/zoology umanitoba.ca/faculties/afs umanitoba.ca/afs umanitoba.ca/faculties/medicine/physiology scrc.umanitoba.ca/Physiology umanitoba.ca/faculties/science/microbiology umanitoba.ca/science/microbiology umanitoba.ca/faculties/science/botany umanitoba.ca/science/botany	145	706700	4962	637
McGill University	mcgill.ca	medicine.mcgill.ca/physio mcgill.ca/parasitology mcgill.ca/microimm mcgill.ca/biology medicine.mcgill.ca/anatomy mcgill.ca/plant	131	1053870	47325	2630
McMaster University	mcmaster.ca	science.mcmaster.ca/biology	31	113850	8022	515
Memorial University of Newfoundland	mun.ca	mun.ca/biology	40	73000	2057	289
University of Montréal	umontreal.ca	medvet.umontreal.ca/biomed medvet.umontreal.ca/CRRA physio.umontreal.ca medvet.umontreal.ca/pathologie_microbiologie bio.umontreal.ca	127	321760	17363	1445
Mount Allison University	mta.ca	mta.ca/faculty/science/bio	16	72000	284	37

University of New Brunswick	unb.ca unbsj.ca	unbsj.ca/sase/biology unb.ca/fredericton/science/biology	47	328917	1021	138
University of Northern British Columbia	unbc.ca	unbc.ca/biology	24	60000	411	54
Nova Scotia Agricultral College	nsac.ns.ca	nsac.ns.ca/pas	24	37400	273	59
Okanagan University College	ouc.bc.ca	ouc.bc.ca/boil	21	56000	18	9
University of Ottawa	uottawa.ca	medicine.uottawa.ca/microbio/bmi bio.uottawa.ca	57	329520	6349	534
Québec à Montréal	uqam.ca	unites.uqam.ca/dsbio	50	493908	2444	293
Québec à Rimouski	uqar.qc.ca uqar.uquebec.ca	uqar.uquebec.ca/bcss uqar.qc.ca/bcss	19	20000	593	78
Queen's University	queensu.ca	biology.queensu.ca microimm.queensu.ca	50	492592	14110	1133
University of Regina	uregina.ca	uregina.ca/biology	9	73577	713	78
Ryerson University	ryerson.ca	ryerson.ca/~cab ryerson.ca/cab ryerson.ca/~cbce	25	57983	110	17
Saint Mary's University	stmarys.ca	stmarys.ca/academic/science/biology	11	38000	113	27
Unviersity of Saskatchewan	usask.ca	usask.ca/wcvm/vbs usask.ca/anatomy usask.ca/biology usask.ca/medicine/microbio usask.ca/agriculture/plantsci	95	254342	5472	653
Université de Sherbrooke	usherbrooke.ca usherb.ca	usherbrooke.ca/anatbiocell usherbrooke.ca/biologie	25	300869	4750	392
Simon Fraser University	sfu.ca	sfu.ca/biology sfu.ca/mbb	59	319405	6146	686
St. Francis Xavier University	stfx.ca	stfx.ca/academic/biology	12	29500	423	39

University of Toronto(remove the grant and staff no. for physiology)	utoronto.ca	zoo.utoronto.ca utoronto.ca/~lifesci utoronto.ca/nutrisci immune.med.utoronto.ca botany.utoronto.ca utoronto.ca/medicalgenetics	289	957959	34807	1690
Trent University	trentu.ca	trentu.ca/biology	22	48000	808	85
University of Victoria	uvic.ca	uvic.ca/biochem uvic.ca/biology	46	528357	4057	365
University of Waterloo	uwaterloo.ca	uwaterloo.ca/biology	33	285065	5923	588
University of Western Ontario	uwo.ca	uwo.ca/biology mni.uwo.ca physpharm.fmd.uwo.ca physpharm.med.uwo.ca	160	409774	8966	773
York University	yorku.ca	biol.yorku.ca	47	314103	4040	358

Appendix C8 Links at Different Document Levels from SocSciBot for Canadian Biology Departments

University Name	F-inlink	Dir-inlink	Dom-inlink	Dept-inlink	F-outlink	Dir-outlink	Dom-outlink	Dept-outlink	Web pages
Acadia University	11	10	5	5	1	1	1	1	264
University of Alberta	60	49	23	18	12	11	9	8	1486
Brandon University	0	0	0	0	1	1	1	1	20
University of British Columbia	39	33	17	16	36	25	19	18	6869
University of Brock	2	2	1	1	0	0	0	0	73
University of Calgary	7	7	7	7	6	5	2	2	940
Concordia University	1	1	1	1	0	0	0	0	32
Dalhousie University	7	7	7	7	3	3	2	2	472
University of Guelph	59	41	19	17	10	9	8	8	9123
Lakehead University	1	1	1	1	0	0	0	0	14
Laurentian University	3	3	2	2	10	9	4	4	174
Université Laval	1	1	1	1	1	1	1	1	2272
University of Lethbridge	3	3	3	3	0	0	0	0	1003
University of Manitoba	8	8	7	7	21	21	9	8	12757
McGill University	10	9	8	8	21	14	6	6	1771
McMaster University	12	12	6	5	0	0	0	0	899
Memorial University of Newfoundland	1	1	1	1	10	9	4	3	8146
University of Montréal	0	0	0	0	2	2	2	2	579
Mount Allison University	1	1	1	1	0	0	0	0	122
University of New Brunswick	3	3	3	3	1	1	1	1	590
University of Northern British Columbia	0	0	0	0	0	0	0	0	11
Nova Scotia Agricultral College	3	3	2	2	4	4	2	2	1172

Okanagan University College	0	0	0	0	2	2	2	2	74
University of Ottawa	9	8	7	5	6	6	3	2	2678
Québec à Montréal	0	0	0	0	0	0	0	0	174
Québec à Rimouski	0	0	0	0	0	0	0	0	119
Queen's University	25	21	10	10	65	58	27	24	5435
University of Regina	6	6	3	3	6	4	4	4	445
Ryerson University	0	0	0	0	0	0	0	0	104
Saint Mary's University	0	0	0	0	0	0	0	0	75
Unviersity of Saskatchewan	8	8	5	5	6	6	4	4	1429
Université de Sherbrooke	0	0	0	0	0	0	0	0	186
Simon Fraser University	22	13	7	7	8	7	6	5	7264
St. Francis Xavier University	0	0	0	0	0	0	0	0	26
University of Toronto(remove the grant and staff no. for physiology)	40	28	15	14	71	64	40	36	8093
Trent University	4	4	3	3	46	27	14	13	373
University of Victoria	3	3	2	2	0	0	0	0	132
University of Waterloo	8	7	5	5	9	5	2	2	2643
University of Western Ontario	4	4	3	3	12	10	9	9	4849
York University	10	9	8	6	1	1	1	1	724

Appendix C9 Inlinks from Different Domains for Canadian Biology Departments from AltaVista

University Name	External	com	edu	ca	org	net	Web pages
Acadia University	59	9	30	9	2	1	0
University of Alberta	4525	936	901	852	486	210	6154
Brandon University	8	4	3	0	1	0	20
University of British Columbia	2331	515	680	446	242	93	3834
University of Brock	14	3	7	1	2	0	39
University of Calgary	249	85	0	35	13	19	402
Concordia University	20	2	11	5	0	0	5
Dalhousie University	332	96	39	34	20	9	330
University of Guelph	91	22	25	21	8	5	62
Lakehead University	13	4	15	0	0	0	14
Laurentian University	38	7	21	0	2	1	134
Université Laval	176	43	62	14	10	4	265
University of Lethbridge	156	14	16	72	6	14	484
University of Manitoba	666	200	153	77	63	22	3338
McGill University	468	89	87	114	57	24	418
McMaster University	1003	256	176	177	110	63	334
Memorial University of Newfoundland	355	59	36	102	37	16	5021
University of Montréal	56	13	26	6	5	0	137
Mount Allison University	72	18	21	20	3	1	126
University of New Brunswick	16	1	11	2	0	1	306
University of Northern British Columbia	2	0	0	2	0	0	4
Nova Scotia Agricultral College	73	17	29	7	1	2	694
Okanagan University College	0	0	0	0	0	0	0
University of Ottawa	2000	419	177	629	146	68	1090
Québec à Montréal	9	0	7	1	0	0	6
Québec à Rimouski	7	0	6	0	0	1	13
Queen's University	1166	250	232	199	100	65	1804

University of Regina	65	10	21	8	3	2	311
Ryerson University	19	5	7	0	0	0	60
Saint Mary's University	20	1	7	9	0	2	31
Unviersity of Saskatchewan	576	141	123	77	65	20	1520
Université de Sherbrooke	2	0	1	0	0	0	46
Simon Fraser University	1240	810	128	80	33	16	2337
St. Francis Xavier University	28	9	6	8	2	1	23
University of Toronto(remove the grant and staff no. for physiology)	2507	608	302	617	413	116	5739
Trent University	114	26	18	14	16	5	368
University of Victoria	218	58	45	24	14	10	117
University of Waterloo	143	32	37	25	8	12	202
University of Western Ontario	351	105	39	62	38	16	928
York University	199	42	59	29	12	9	715

Appendix D Data for the UK Physics, Chemistry and Biology Departments

Appendix D1 Non-web Data for the UK Physics Departments

University Name	University domain	Department domain	Staff members	RAE averages	Citation counts	Publications
University of Bath	Bath.ac.uk	bath.ac.uk/physics	24	5	2354	297
University of Birmingham	bham.ac.uk birmingham.ac.uk	ph.bham.ac.uk	62.36	5.064143682	12897	1126
University of Bristol	bris.ac.uk bristol.ac.uk	phy.bris.ac.uk	49	5.755102041	8840	558
University of Cambridge	cam.ac.uk Cambridge.ac.uk	phy.cam.ac.uk mrao.cam.ac.uk ast.cam.ac.uk	138.88	7.000288018	51613	2779
University of Central Lancashire	uclan.ac.uk	uclan.ac.uk/facs/science/physastr	19.15	4.219321149	934	125
University of Durham	dur.ac.uk durham.ac.uk	dur.ac.uk/Physics durham.ac.uk/Physics	55	6	22161	1299
University of Exeter	ex.ac.uk exeter.ac.uk	newton.ex.ac.uk	21	5.714285714	2205	286
University of Hertfordshire	herts.ac.uk hertfordshire.ac.uk	herts.ac.uk/natsci/physics hertfordshire.ac.uk/natsci/physics strc.herts.ac.uk/astro strc.herts.ac.uk/ls strc.herts.ac.uk/tp	26	4.230769231	1918	180
Imperial College of Science, Technology and Medicine	imperial.ac.uk ic.ac.uk	imperial.ac.uk/physics ic.ac.uk/physics	102.83	6.795682194	934	119
Keele University	keele.ac.uk	phys.keele.ac.uk	20.5	2.414634146	1657	211
University of Kent at Canterbury	kent.ac.uk ukc.ac.uk	kent.ac.uk/physical-sciences ukc.ac.uk/physical-sciences	23.53	2.192945176	1125	152
King's College London	kcl.ac.uk	kcl.ac.uk/kis/schools/phys_eng/physics ph.kcl.ac.uk	18.5	4.459459459	3356	360

Lancaster University	lancs.ac.uk lancaster.ac.uk	lancs.ac.uk/depts/physics lancaster.ac.uk/depts/physics	21.83	6.678882272	5388	456
University of Leeds	leeds.ac.uk	.physics.leeds.ac.uk	39.75	5.969811321	5046	612
University of Leicester	le.ac.uk leicester.ac.uk	le.ac.uk/physics leicester.ac.uk/physics	43	5.720930233	4651	487
University of Liverpool	liv.ac.uk liverpool.ac.uk	physics.liv.ac.uk ph.liv.ac.uk	36.2	5.834254144	6993	680
Liverpool John Moores University	livjm.ac.uk	astro.livjm.ac.uk	16	5	774	76
Loughborough University	lboro.ac.uk loughborough.ac.uk	lboro.ac.uk/departments/ph loughborough.ac.uk/departments/ph	10	3.5	1807	321
University of Manchester	man.ac.uk manchester.ac.ukmcc.ac.uk	physics.man.ac.uk	61	5.803278689	4676	676
University of Manchester Institute of Science & Technology	umist.ac.uk	umist.ac.uk/departments/physics	32.33	4.846891432	3051	518
University of Newcastle	ncl.ac.uk newcastle.ac.uk	phys.ncl.ac.uk phys.newcastle.ac.uk	16.68	4.340527578	1832	266
University of Nottingham	nottingham.ac.uk nott.ac.uk	nottingham.ac.uk/physics nott.ac.uk/physics	32.8	5.817073171	4667	670
Open University	open.ac.uk	physics.open.ac.uk	29.7	2.02020202	1493	196
University of Oxford	ox.ac.uk oxford.ac.uk	physics.ox.ac.uk	159.6	6.824561404	8557	835
Queen Mary, University of London	qmul.ac.uk qmw.ac.uk	ph.qmul.ac.uk ph.qmw.ac.uk maths.qmul.ac.uk/Astronomy maths.qmw.ac.uk/Astronomy	52	5.653846154	9623	766
University of Reading	rdg.ac.uk reading.ac.uk	rdg.ac.uk/Physics reading.ac.uk/Physics	20.3	4.507389163	725	70
Royal Holloway, University of London	rhul.ac.uk rhbnc.ac.uk	ph.rhul.ac.uk ph.rhbnc.ac.uk	22.73	5.5433348	992	111

University of Sheffield	shef.ac.uk Sheffield.ac.uk	shef.ac.uk/physics sheffield.ac.uk/physics shef.ac.uk/~mpce sheffield.ac.uk/~mpce	30	5.6	5573	456
University of Southampton	soton.ac.uk southampton.ac.uk	.phys.soton.ac.uk	34.88	6.152522936	7482	746
University of Surrey	surrey.ac.uk	ph.surrey.ac.uk	28	6	1242	199
University of Sussex	sussex.ac.uk susx.ac.uk	sussex.ac.uk/physics susx.ac.uk/physics	22.67	5.734450816	3557	217
University College London	ucl.ac.uk	phys.ucl.ac.uk	92.35	5.480238224	7121	788
University of Warwick	Warwick.ac.uk	phys.warwick.ac.uk	31.33	5.809128631	3558	542
University of York	york.ac.uk	york.ac.uk/depts/phys	24	4.583333333	938	194
University of Edinburgh	ed.ac.uk edinburgh.ac.uk	ph.ed.ac.uk epcc.ed.ac.uk epcc.edinburgh.ac.uk	72.8	5.340659341	7949	470
University of Glasgow	gla.ac.uk Glasgow.ac.uk	physics.gla.ac.uk	45.99	5.305501196	4929	550
Heriot-Watt University	hw.ac.uk heriot-watt.ac.uk	phy.hw.ac.uk	32.25	4.846511628	1282	237
University of Paisley	paisley.ac.uk	eep.paisley.ac.uk/physics	8	2.625	104	18
University of St Andrews	st-andrews.ac.uk st-and.ac.uk	st-andrews.ac.uk/physics st-and.ac.uk/physics	30	5.4	3326	368
University of Strathclyde	strath.ac.uk strathclyde.ac.uk	phys.strath.ac.uk	48	4.760416667	2409	307
University of Wales, Swansea	swansea.ac.uk swan.ac.uk	physics.swansea.ac.uk physics.swan.ac.uk python.swan.ac.uk python.swansea.ac.uk	11.63	6.00171969	1914	229
University of Wales, Aberystwyth	aber.ac.uk aberystwyth.ac.uk	aber.ac.uk/physics aberystwyth.ac.uk/physics	15.33	5.003261579	225	45
Cardiff University	cf.ac.uk cardiff.ac.uk	astro.cf.ac.uk astro.cardiff.ac.uk	24.03	5.751144403	262	47

The Queen's University of Belfast	qub.ac.uk	physics.qub.ac.uk	57	5.789473684	4836	632

Appendix D2 Links at Different Document Levels from SocSciBot for the UK Physics Departments

University Name	F-inlink	Dir-inlink	Dom-inlink	Dept-inlink	F-outlink	Dir-outlink	Dom-outlink	Dept-outlink	Web pages
University of Bath	1	1	1	1	4	1	1	1	127
University of Birmingham	65	50	22	12	26	23	21	15	2965
University of Bristol	31	30	20	13	6	5	5	4	2983
University of Cambridge	109	103	45	17	294	133	64	21	54624
University of Central Lancashire	0	0	0	0	1	1	1	1	172
University of Durham	10	9	9	7	3	3	3	3	515
University of Exeter	22	16	12	11	30	25	9	8	3830
University of Hertfordshire	9	1	1	1	4	2	2	2	735
Imperial College of Science, Technology and Medicine	0	0	0	0	0	0	0	0	114
Keele University	1	1	1	1	46	45	31	23	357
University of Kent at Canterbury	1	1	1	1	1	1	1	1	594
King's College London	18	18	14	10	18	17	13	7	486
Lancaster University	3	3	3	3	5	5	5	5	1581
University of Leeds	10	3	1	1	0	0	0	0	267
University of Leicester	6	4	4	3	1	1	1	1	905
University of Liverpool	42	38	19	11	67	57	20	11	12046
Liverpool John Moores University	6	4	4	2	10	8	6	5	1178
Loughborough University	5	4	3	3	5	5	4	4	186
University of Manchester	0	0	0	0	0	0	0	0	77
University of Manchester Institute of Science & Technology	0	0	0	0	1	1	1	1	98
University of Newcastle	10	8	6	6	3	3	3	3	678
University of Nottingham	13	6	5	4	7	5	4	4	405

Open University	38	10	8	7	2	2	2	2	499
University of Oxford	207	144	68	19	100	93	46	19	20781
Queen Mary, University of London	19	11	7	5	28	21	12	7	2796
University of Reading	3	3	3	2	0	0	0	0	52
Royal Holloway, University of London	2	2	2	1	24	16	15	11	2398
University of Sheffield	0	0	0	0	42	24	15	10	3806
University of Southampton	46	43	26	15	40	31	16	9	1213
University of Surrey	20	19	14	12	57	53	31	19	3686
University of Sussex	10	9	7	7	0	0	0	0	1586
University College London	30	16	13	10	2	2	2	2	1069
University of Warwick	9	9	6	5	13	11	8	5	1712
University of York	14	12	7	5	10	3	1	1	362
University of Edinburgh	113	84	28	17	76	71	23	13	23732
University of Glasgow	108	58	18	12	38	37	10	7	1703
Heriot-Watt University	21	21	10	8	8	8	7	5	2548
University of Paisley	0	0	0	0	0	0	0	0	14
University of St Andrews	1	1	1	1	11	11	2	2	284
University of Strathclyde	13	10	8	7	50	19	12	7	3936
University of Wales, Swansea	0	0	0	0	4	2	2	2	161
University of Wales, Aberystwyth	0	0	0	0	0	0	0	0	171
Cardiff University	53	23	13	8	31	29	10	6	3550
The Queen's University of Belfast	0	0	0	0	1	1	1	1	280

Appendix D3 Inlinks from Different Domains for the UK Physics Departments from AltaVista

University Name	External	com	edu	uk	org	net	co.uk	ac.uk	Web pages
University of Bath	54	12	0	19	5	3	6	8	116
University of Birmingham	1089	199	78	1	56	33	33	1	372
University of Bristol	2052	359	213	529	118	78	38	454	2352
University of Cambridge	20590	2761	2309	3261	1360	824	678	2186	18838
University of Central Lancashire	96	17	6	28	6	11	2	25	139
University of Durham	162	12	13	58	9	3	12	39	370
University of Exeter	4840	520	307	352	121	118	80	262	5994
University of Hertfordshire	81	5	7	27	5	3	5	20	284
Imperial College of Science, Technology and Medicine	3	1	0	1	0	0	0	1	22
Keele University	130	9	3	85	1	4	2	79	222
University of Kent at Canterbury	152	20	8	77	20	8	16	47	674
King's College London	1537	186	214	267	60	48	18	231	376
Lancaster University	111	14	12	26	6	5	7	19	595
University of Leeds	77	7	4	36	3	3	3	27	361
University of Leicester	245	29	0	82	17	9	9	63	475
University of Liverpool	673	42	25	228	10	11	28	169	6631
Liverpool John Moores University	231	22	15	109	35	6	18	41	1353
Loughborough University	53	8	3	20	3	2	5	11	176
University of Manchester	116	9	5	90	1	2	2	88	88
University of Manchester Institute of Science & Technology	15	3	0	6	2	0	1	4	103
University of Newcastle	502	33	15	373	6	9	15	356	389
University of Nottingham	174	40	10	32	16	7	8	21	401
Open University	487	35	38	175	30	17	13	145	858

University of Oxford	5511	901	514	1417	406	232	163	1216	27312
Queen Mary, University of London	2113	372	172	643	113	76	87	512	13758
University of Reading	65	8	1	25	0	4	5	18	66
Royal Holloway, University of London	551	60	29	278	8	16	18	244	2180
University of Sheffield	26	9	7	29	1	0	0	19	1665
University of Southampton	1101	143	168	371	37	39	18	331	826
University of Surrey	1056	187	79	331	98	30	31	271	2039
University of Sussex	112	11	13	32	2	3	7	23	1807
University College London	683	67	43	363	25	22	21	333	858
University of Warwick	272	9	17	179	11	5	12	154	582
University of York	167	16	10	75	6	10	5	68	196
University of Edinburgh	5266	642	491	1098	418	324	126	894	15933
University of Glasgow	1274	149	108	497	54	30	30	455	1531
Heriot-Watt University	1250	336	68	392	45	72	161	189	1123
University of Paisley	6	0	0	6	0	0	1	5	15
University of St Andrews	29	7	4	6	4	1	0	6	13
University of Strathclyde	601	98	58	178	13	15	30	139	2952
University of Wales, Swansea	379	73	15	148	13	24	25	105	142
University of Wales, Aberystwyth	38	4	1	18	6	3	7	9	40
Cardiff University	1618	424	189	329	106	78	91	211	1751
The Queen's University of Belfast	1	0	0	1	0	0	0	1	15

Appendix D4 Non-web Data for the UK Chemistry Departments

University Name	University domain	Department domain	Staff members	RAE averages	Citation counts	Publications
University of Bath	bath.ac.uk	bath.ac.uk/chemistry	25.45	4.805500982	4239	392
University of Birmingham	bham.ac.uk birmingham.ac.uk	chem.bham.ac.uk	42.05	4.887039239	6447	595
University of Bristol	bris.ac.uk bristol.ac.uk	bris.ac.uk/Depts/Chemistry bristol.ac.uk/Depts/Chemistry chm bris.ac.uk chm.bristol.ac.uk	44	6.522727273	7167	634
University of Cambridge	cam.ac.uk cambridge.ac.uk	ch.cam.ac.uk	68	7	15052	1166
University of Durham	dur.ac.uk durham.ac.uk	dur.ac.uk/chemistry durham.ac.uk/chemistry	36.5	7	6763	662
University of East Anglia	uea.ac.uk	uea.ac.uk/che	25.5	5.176470588	2224	266
University of Exeter	ex.ac.uk exeter.ac.uk	ex.ac.uk/chemweb exeter.ac.uk/chemweb	28.4	4.471830986	3181	319
University of Huddersfield	hud.ac.uk huddersfield.ac.uk	hud.ac.uk/schools/applied_sciences/chemistry huddersfield.ac.uk/schools/applied_sciences/chemistry hud.ac.uk/sas huddersfield.ac.uk/sas	18.5	1.837837838	631	89
University of Hull	hull.ac.uk	hull.ac.uk/chemistry hull.ac.uk/Hull/Chem_Web	33.25	3.94887218	3280	341
Imperial College of Science, Technology and Medicine	ic.ac.uk imperial.ac.uk	ch.ic.ac.uk ch.imperial.ac.uk	46.95	6.404685836	1258	123
Keele University	keele.ac.uk	keele.ac.uk/depts/ch	15	3	854	96
King's College London	kcl.ac.uk	ch.kcl.ac.uk	24.4	4.508196721	2830	283
University of Leeds	leeds.ac.uk	chem.leeds.ac.uk colour.leeds.ac.uk	53.1	5.548022599	7188	777

Loughborough University	lboro.ac.uk loughborough.ac.uk	lboro.ac.uk/departments/cm loughborough.ac.uk/departments/cm	27	4.259259259	3234	391
University of Manchester	man.ac.uk manchester.ac.uk mcc.ac.uk	ch.man.ac.uk	36.2	5.337016575	5578	564
University of Manchester Institute of Science & Technology	umist.ac.uk	umist.ac.uk/departments/chemistry umist.ac.uk/chemistry	32.6	4.539877301	4954	582
University of Newcastle	ncl.ac.uk Newcastle.ac.uk	ncl.ac.uk/chemistry newcastle.ac.uk/chemistry	30	3.833333333	3390	383
University of Northumbria at Newcastle	northumbria.ac.uk unn.ac.uk	northumbria.ac.uk/faculties/est/ams unn.ac.uk/faculties/est/ams unn.ac.uk/academic/est/cls northumbria.ac.uk/academic/est/cls northumbria.ac.uk/sd/academic/sas	15	2.666666667	704	64
University of Nottingham	nottingham.ac.uk nott.ac.uk	nottingham.ac.uk/chemistry nott.ac.uk/chemistry	41.9	5.856801909	9231	716
Nottingham Trent University	ntu.ac.uk	science.ntu.ac.uk/chph	23	1.304347826	750	109
Open University	open.ac.uk	open.ac.uk/science/chemistry	22.2	2.027027027	546	96
University of Oxford	ox.ac.uk oxford.ac.uk	chem.ox.ac.uk	76.75	6.727035831	12224	896
Queen Mary, University of London	qmul.ac.uk qmw.ac.uk	chem.qmul.ac.uk chem.qmw.ac.uk	22.5	2.733333333	1419	198
University of Reading	rdg.ac.uk reading.ac.uk	chem.rdg.ac.uk chem.reading.ac.uk	27	4.814814815	2708	327
University of Sheffield	shef.ac.uk sheffield.ac.uk	shef.ac.uk/chemistry sheffield.ac.uk/chemistry	36	5	4281	353
University of Southampton	soton.ac.uk southampton.ac.uk	chem.soton.ac.uk	39.45	5.771863118	5854	592
University of Surrey	surrey.ac.uk	surrey.ac.uk/Chemistry	21.4	2.859813084	582	129
University of Sussex	sussex.ac.uk susx.ac.uk	sussex.ac.uk/chemistry susx.ac.uk/chemistry	33	6	4339	476

University College London	ucl.ac.uk	chem.ucl.ac.uk	34.75	6.397122302	5561	518
University of Warwick	warwick.ac.uk	warwick.ac.uk/fac/sci/chemistry	25	5.76	5430	471
University of York	york.ac.uk	york.ac.uk/depts/chem. york.ac.uk/res/gcg	44.55	5.865319865	4019	351
University of Aberdeen	abdn.ac.uk aberdeen.ac.uk	abdn.ac.uk/chemistry aberdeen.ac.uk/chemistry	18.1	2.187845304	1660	178
University of Edinburgh	ed.ac.uk edinburgh.ac.uk	chem.ed.ac.uk chem.edinburgh.ac.uk	44	5.863636364	2227	288
University of Glasgow	gla.ac.uk Glasgow.ac.uk	chem.gla.ac.uk	38.86	3.795676788	3432	328
Heriot-Watt University	hw.ac.uk heriot-watt.ac.uk	hw.ac.uk/cheWWW heriot-watt.ac.uk/cheWWW	26.6	4.812030075	1216	168
University of St Andrews	st-and.ac.uk st-andrews.ac.uk	ch-www.st-and.ac.uk ch-www.st-andrews.ac.uk	31	4.451612903	3270	375
University of Strathclyde	strath.ac.uk strathclyde.ac.uk	strath.ac.uk/Departments/Chemistry strathclyde.ac.uk/Departments/Chemistry strath.ac.uk/chemistry strathclyde.ac.uk/chemistry chem.strath.ac.uk	43.84	4.90419708	4008	276
University of Wales, Bangor[2]	bangor.ac.uk	bangor.ac.uk/ch	12.3	3	966	116
University of Wales, Swansea	swan.ac.uk Swansea.ac.uk	swan.ac.uk/chemistry swansea.ac.uk/chemistry	12	5	1318	202
Cardiff University	cf.ac.uk cardiff.ac.uk	cf.ac.uk/chemy cardiff.ac.uk/chemy	34.5	5	1758	235
The Queen's University of Belfast	qub.ac.uk	ch.qub.ac.uk	28	5	449	62

Appendix D5 Links at Different Document Levels from SocSciBot for the UK Chemistry Departments

University Name	F-inlink	Dir-inlink	Dom-inlink	Dept-inlink	F-outlink	Dir-outlink	Dom-outlink	Dept-outlink	Web pages
University of Bath	8	8	6	6	1	1	1	1	102
University of Birmingham	135	21	7	6	2	2	2	2	330
University of Bristol	273	98	37	22	162	112	41	23	10859
University of Cambridge	344	127	36	20	2570	717	85	38	10081
University of Durham	113	57	7	6	11	11	10	10	314
University of East Anglia	13	10	5	5	1	1	1	1	24
University of Exeter	87	25	8	8	11	11	8	6	438
University of Huddersfield	7	4	2	2	1	1	1	1	226
University of Hull	61	13	6	5	0	0	0	0	66
Imperial College of Science, Technology and Medicine	312	215	34	21	241	149	27	20	18423
Keele University	4	4	3	2	9	9	4	4	490
King's College London	71	16	8	6	16	13	8	7	945
University of Leeds	219	78	16	11	23	20	10	9	2110
Loughborough University	3	3	3	3	2	2	1	1	225
University of Manchester	107	28	17	11	22	20	11	9	4541
University of Manchester Institute of Science & Technology	5	5	4	3	6	6	6	5	315
University of Newcastle	131	23	7	7	1	1	1	1	228
University of Northumbria at Newcastle	5	4	2	2	0	0	0	0	313
University of Nottingham	96	57	9	7	2	1	1	1	471
Nottingham Trent University	81	19	9	7	6	6	5	5	426
Open University	2	2	1	1	2	2	2	2	64
University of Oxford	294	105	23	15	136	132	14	12	9830
Queen Mary, University of London	5	5	4	4	31	18	10	10	7434

University of Reading	96	26	8	8	7	7	4	4	239
University of Sheffield	223	175	21	14	35	28	14	14	852
University of Southampton	10	10	7	6	0	0	0	0	700
University of Surrey	5	5	3	3	6	6	1	1	244
University of Sussex	5	4	4	4	10	9	6	3	184
University College London	193	98	13	9	64	60	36	34	1066
University of Warwick	84	48	7	5	0	0	0	0	29
University of York	166	25	14	12	21	15	9	8	1169
University of Aberdeen	64	12	6	5	284	72	12	11	1038
University of Edinburgh	104	14	11	9	78	67	24	17	1946
University of Glasgow	122	73	17	14	53	47	24	21	2688
Heriot-Watt University	82	21	10	9	0	0	0	0	103
University of St Andrews	90	53	8	7	11	7	7	7	506
University of Strathclyde	19	17	10	8	22	20	16	12	318
University of Wales, Bangor[2]	56	28	7	6	11	9	7	7	152
University of Wales, Swansea	3	3	1	1	0	0	0	0	662
Cardiff University	96	15	10	10	2	2	2	2	267
The Queen's University of Belfast	70	34	3	2	4	4	3	3	98

Appendix D6 Inlinks from Different Domains for the UK Chemistry Departments from AltaVista

University Name	External	com	edu	uk	org	net	co.uk	ac.uk	Web pages
University of Bath	40	4	5	23	2	2	7	13	141
University of Birmingham	185	6	6	142	3	5	3	133	248
University of Bristol	4681	1040	786	735	384	230	211	465	223
University of Cambridge	7164	1300	1034	1386	432	334	253	1039	7401
University of Durham	46	5	6	23	5	1	2	21	309
University of East Anglia	98	10	3	62	7	2	8	44	150
University of Exeter	135	17	10	42	4	2	6	35	341
University of Huddersfield	92	17	1	28	11	2	5	16	202
University of Hull	147	12	16	57	10	5	4	46	557
Imperial College of Science, Technology and Medicine	7139	926	1126	1637	391	258	141	1452	8245
Keele University	172	21	27	61	17	4	6	36	473
King's College London	342	75	31	98	23	9	5	90	667
University of Leeds	2687	503	347	463	176	79	79	354	1841
Loughborough University	76	17	2	22	2	4	3	15	198
University of Manchester	1131	335	84	291	42	41	71	212	2063
University of Manchester Institute of Science & Technology	46	7	1	28	1	1	2	24	310
University of Newcastle	78	14	3	26	5	0	3	23	294
University of Northumbria at Newcastle	23	4	1	16	0	0	2	12	3
University of Nottingham	176	34	11	73	15	4	6	66	198
Nottingham Trent University	141	25	5	66	7	4	13	45	485
Open University	18	1	0	9	4	0	3	5	83
University of Oxford	9447	2011	699	5003	339	241	287	4636	7602
Queen Mary, University of London	5323	1014	948	1093	292	135	35	1030	4314
University of Reading	672	150	44	272	19	24	60	205	150

University of Sheffield	7983	3081	1349	376	427	222	64	252	549
University of Southampton	199	5	2	168	6	0	1	164	712
University of Surrey	72	10	2	28	1	3	1	26	115
University of Sussex	40	2	2	29	2	0	4	20	329
University College London	710	115	40	234	85	25	29	192	788
University of Warwick	230	92	15	69	7	2	12	46	868
University of York	698	130	116	168	28	14	20	133	833
University of Aberdeen	72	11	4	26	2	3	1	25	365
University of Edinburgh	543	78	35	216	21	30	30	181	1147
University of Glasgow	1309	194	158	411	72	34	78	316	1398
Heriot-Watt University	126	11	14	41	3	2	2	38	225
University of St Andrews	236	39	10	142	18	5	8	129	291
University of Strathclyde	346	91	13	85	21	8	18	63	324
University of Wales, Bangor[2]	113	16	6	37	8	5	6	27	105
University of Wales, Swansea	83	15	6	26	8	3	10	15	655
Cardiff University	97	12	5	48	4	5	6	41	236
The Queen's University of Belfast	447	85	36	97	33	17	10	86	124

Appendix D7 Non-web Data for the UK Biology Departments

University Name	University domain	Department domain	Staff members	RAE averages	Citation counts	Publications
University of Bath	bath.ac.uk	bath.ac.uk/bio-sci	47.45	5.443625	9021	617
Birkbeck College	bbk.ac.uk birkbeck.ac.uk	cryst.bbk.ac.uk bbk.ac.uk/bcs	40.7	4.289926	5930	353
University of Birmingham	bham.ac.uk birmingham.ac.uk	biosciences.bham.ac.uk	61.96	4.912847	25699	1649
University of Bristol	bris.ac.uk bristol.ac.uk	bio.bris.ac.uk bio.bristol.ac.uk bch.bris.ac.uk bch.bristol.ac.uk	69.43	6.39205	18918	1282
Brunel University	brunel.ac.uk	brunel.ac.uk/depts/bio	25.85	3.531915	1804	104
University of Cambridge	cam.ac.uk cambridge.ac.uk	bioc.cam.ac.uk gen.cam.ac.uk plantsci.cam.ac.uk zoo.cam.ac.uk biot.cam.ac.uk	162.66	6.297184	34575	1969
University of Central Lancashire	uclan.ac.uk	uclan.ac.uk/biology uclan.ac.uk/facs/science/biology	17.75	3.211268	662	117
Cranfield University	cranfield.ac.uk	cranfield.ac.uk/ibst cranfield.ac.uk/biotech	11.5	5	651	91
University of Durham	dur.ac.uk durham.ac.uk	dur.ac.uk/biological.sciences durham.ac.uk/biological.sciences	34	5.117647	3975	395
University of East Anglia	uea.ac.uk	uea.ac.uk/bio	58.23	5.641422	10868	610
University of East London	uel.ac.uk	uel.ac.uk/hab	9.4	0.851064	569	64
University of Essex	essex.ac.uk sx.ac.uk	essex.ac.uk/bs sx.ac.uk/bs	34.65	4.020202	4362	267
University of Exeter	ex.ac.uk exeter.ac.uk	ex.ac.uk/biology	23.48	4.55707	1453	174

University of Hull	hull.ac.uk	hull.ac.uk/biosci	16.17	2.937539	1989	233
Imperial College of Science, Technology and Medicine	ic.ac.uk imperial.ac.uk	bio.ic.ac.uk	82.9	6.3076	20915	1400
Keele University	keele.ac.uk	keele.ac.uk/depts/bi keele.ac.uk/depts/aep keele.ac.uk/depts/co	27	3.611111	1764	207
University of Kent at Canterbury	kent.ac.uk ukc.ac.uk	kent.ac.uk/bio ukc.ac.uk/bio wwwbio.ukc.ac.uk	23.6	4.775424	2527	195
King's College London	kcl.ac.uk	kcl.ac.uk/kis/schools/life_sciences kcl.ac.uk/depsta/biomedical/randall	65.5	3.045802	5025	453
Lancaster University	lancs.ac.uk lancaster.ac.uk	bssv01.lancs.ac.uk/bs	31.36	4.550383	3133	237
University of Leeds	leeds.ac.uk	biology.leeds.ac.uk	81.1	5.030826	13743	1205
University of Leicester	le.ac.uk leicester.ac.uk	le.ac.uk/biochem leicester.ac.uk/biochem le.ac.uk/biology leicester.ac.uk/biology le.ac.uk/genetics leicester.ac.uk/genetics le.ac.uk/bs	61.63	5.356158	10542	700
University of Luton	luton.ac.uk	luton.ac.uk/depts/sport_exercise_sci luton.ac.uk/lirans	20.8	0.942308	182	30
University of Manchester	man.ac.uk manchester.ac.uk mcc.ac.uk	biomed.man.ac.uk	82.03	5.802755	21867	1451
University of Manchester Institute of Science & Technology	umist.ac.uk	bi.umist.ac.uk	33	6	4223	374
Manchester Metropolitan University	mmu.ac.uk	sci-eng.mmu.ac.uk/biology	26.5	1.056604	488	73

University of Newcastle	ncl.ac.uk Newcastle.ac.uk	medical.faculty.ncl.ac.uk/biomed/sbg medical.faculty.newcastle.ac.uk/biomed/sbg ncl.ac.uk/ihg newcastle.ac.uk/ihg	27.62	5.984794	19000	1320
University of Northumbria at Newcastle	northumbria.ac.uk	online.northumbria.ac.uk/faculties/est/ams unn.ac.uk/academic/est/cls	16.2	0.617284	346	43
University of Nottingham	nottingham.ac.uk nott.ac.uk	vsb.life.nottingham.ac.uk/life-env horus.cs.nott.ac.uk/life-env	50.6	3.98419	8660	805
Open University	open.ac.uk	open.ac.uk/science/biosci	29.1	2.57732	1554	179
University of Oxford	ox.ac.uk oxford.ac.uk	storage.plants.ox.ac.uk plants.ox.ac.uk zoo.ox.ac.uk bioch.ox.ac.uk	172.37	5.703429	34363	1975
Oxford Brookes University	brookes.ac.uk oxford-brookes.ac.uk	brookes.ac.uk/schools/bms oxford-brookes.ac.uk/schools/bms brookes.ac.uk/bms oxford-brookes.ac.uk/bms	16.7	2.91018	2271	173
University of Plymouth	plym.ac.uk plymouth.ac.uk	science.plym.ac.uk/departments/biology science.plymouth.ac.uk/departments/biology	30.6	0.980392	2069	287
Queen Mary, University of London	qmw.ac.uk qmul.ac.uk	biology.qmw.ac.uk biology.qmul.ac.uk	35	4.571429	5224	428
University of Reading	rdg.ac.uk reading.ac.uk	ams.rdg.ac.uk ams.reading.ac.uk	39	4.423077	3374	219
Royal Holloway, University of London	rhul.ac.uk rhbnc.ac.uk	rhul.ac.uk/biological-sciences rhbnc.ac.uk/biological-sciences	33.65	4.10104	3408	286
University of Sheffield	shef.ac.uk sheffield.ac.uk	shef.ac.uk/aps sheffield.ac.uk/aps shef.ac.uk/mbb sheffield.ac.uk/mbb	58.9	6.049236	11658	824
University of Southampton	soton.ac.uk southampton.ac.uk	biology.soton.ac.uk sobs.soton.ac.uk sbs.soton.ac.uk	38.71	5.536037	6450	630

University of Sussex	susx.ac.uk sussex.ac.uk	biols.susx.ac.uk biols.sussex.ac.uk susx.ac.uk/biology sussex.ac.uk/biology susx.ac.uk/biochemistry sussex.ac.uk/biochemistry susx.ac.uk/psychology sussex.ac.uk/psychology	81.13	5.664982	5927	439
University College London	ucl.ac.uk	biochem.ucl.ac.uk bioinf.cs.ucl.ac.uk ucl.ac.uk/biology earthsciences.ucl.ac.uk	93.35	5.553294	36560	1684
University of Warwick	warwick.ac.uk	bio.warwick.ac.uk	59.98	5.700233	8389	575
University of Westminster	wmin.ac.uk westminster.ac.uk	wmin.ac.uk/biosciences westminster.ac.uk/biosciences wmin.ac.uk/biosciences/Oldschool westminster.ac.uk/biosciences/Oldschool	17.9	2.592179	11	5
University of York	york.ac.uk	york.ac.uk/depts/biol	51.94	5.544859	6433	419
University of Aberdeen	abdn.ac.uk aberdeen.ac.uk	abdn.ac.uk/biologicalsci aberdeen.ac.uk/biologicalsci	29.8	3.825503	2259	246
University of Dundee	dundee.ac.uk	dundee.ac.uk/biocentre dundee.ac.uk/lifesciences dundee.ac.uk/bioscience	83	5.650602	27235	1059
University of Edinburgh	ed.ac.uk edinburgh.ac.uk	biology.ed.ac.uk icapb.ed.ac.uk icmb.ed.ac.uk icmb.edinburgh.ac.uk iscr.ed.ac.uk	161.15	5.702141	24363	1058
University of Glasgow	gla.ac.uk glasgow.ac.uk	gla.ac.uk/ibls gla.ac.uk/Acad/IBLS	93.5	5.294118	14351	1134

University of St Andrews	st-and.ac.uk st-andrews.ac.uk	biology.st-and.ac.uk biology.st-andrews.ac.uk www-sbms.st-and.ac.uk www-sbms.st-andrews.ac.uk	55	4.909091	2337	179
University of Stirling	stir.ac.uk stirling.ac.uk	sbes.stir.ac.uk	28	4.821429	3857	410
University of Wales, Bangor	bangor.ac.uk	biology.bangor.ac.uk	30.33	4.836795	2524	247
University of Wales, Swansea	swan.ac.uk Swansea.ac.uk	swan.ac.uk/biosci swansea.ac.uk/biosci	36	3	2569	261
University of Wales, Aberystwyth	aber.ac.uk aberystwyth.ac.uk	aber.ac.uk/~dbswww aberystwyth.ac.uk/~dbswww aber.ac.uk/biology aberystwyth.ac.uk/biology aber.ac.uk/plantpathol aberystwyth.ac.uk/plantpathol aber.ac.uk/compsci/Research/bio aberystwyth.ac.uk/compsci/Research/bio	28	2.892857	2228	190
Cardiff University	cf.ac.uk cardiff.ac.uk	cf.ac.uk/biosi cardiff.ac.uk/biosi	61.8	5.288026	1885	142
The Queen's University of Belfast	qub.ac.uk	qub.ac.uk/bb	32	4.53125	3486	297

Appendix D8 Links at Different Document Levels from SocSciBot for the UK Biology Departments

University Name	F-inlink	Dir-inlink	Dom-inlink	Dept-inlink	F-outlink	Dir-outlink	Dom-outlink	Dept-outlink	Web pages
University of Bath	3	2	2	2	4	4	4	4	251
Birkbeck College	105	64	36	16	414	306	47	17	16946
University of Birmingham	8	7	6	4	3	3	3	3	319
University of Bristol	36	32	14	10	60	34	18	15	3533
Brunel University	1	1	1	1	11	10	2	2	10533
University of Cambridge	47	37	17	11	1849	1708	56	25	15734
University of Central Lancashire	2	2	2	2	0	0	0	0	121
Cranfield University	0	0	0	0	0	0	0	0	397
University of Durham	0	0	0	0	18	12	4	2	1354
University of East Anglia	7	5	4	4	2	2	2	2	432
University of East London	0	0	0	0	0	0	0	0	183
University of Essex	3	2	2	2	1	1	1	1	794
University of Exeter	2	2	2	2	4	4	2	2	200
University of Hull	0	0	0	0	0	0	0	0	96
Imperial College of Science, Technology and Medicine	39	37	19	14	75	53	25	15	5988
Keele University	1	1	1	1	2	1	1	1	492
University of Kent at Canterbury	0	0	0	0	2	2	2	2	101
King's College London	10	9	7	6	4	4	4	3	1982
Lancaster University	3	3	2	2	3	3	3	3	131
University of Leeds	8	5	3	3	11	9	4	3	884
University of Leicester	28	19	6	5	24	13	10	9	995
University of Luton	0	0	0	0	0	0	0	0	31
University of Manchester	10	8	7	6	1	1	1	1	1385

University of Manchester Institute of Science & Technology	13	12	8	6	4	4	3	3	347
Manchester Metropolitan University	1	1	1	1	4	3	3	3	146
University of Newcastle	0	0	0	0	2	2	2	2	284
University of Northumbria at Newcastle	0	0	0	0	0	0	0	0	243
University of Nottingham	0	0	0	0	0	0	0	0	47
Open University	0	0	0	0	0	0	0	0	81
University of Oxford	60	41	19	11	14	13	9	7	5132
Oxford Brookes University	1	1	1	1	2	2	2	2	1160
University of Plymouth	0	0	0	0	0	0	0	0	59
Queen Mary, University of London	7	4	3	3	1	1	1	1	690
University of Reading	1	1	1	1	9	9	6	6	1179
Royal Holloway, University of London	4	3	3	2	0	0	0	0	10
University of Sheffield	4	4	3	2	12	7	7	7	1835
University of Southampton	2	2	2	2	0	0	0	0	820
University of Sussex	18	17	11	6	7	6	5	5	3603
University College London	2129	1893	38	16	51	21	11	4	2445
University of Warwick	34	28	16	9	20	19	13	9	576
University of Westminster	0	0	0	0	0	0	0	0	174
University of York	13	9	7	5	5	2	2	2	442
University of Aberdeen	0	0	0	0	0	0	0	0	133
University of Dundee	28	10	10	9	5	4	4	4	604
University of Edinburgh	8	7	7	4	3	3	2	2	1455
University of Glasgow	6	6	4	4	4	4	4	4	2524
University of St Andrews	2	1	1	1	20	2	2	2	453
University of Stirling	0	0	0	0	0	0	0	0	211
University of Wales, Bangor	11	6	5	5	1	1	1	1	2619

University of Wales, Swansea	1	1	1	1	0	0	0	0	111
University of Wales, Aberystwyth	0	0	0	0	5	5	3	3	930
Cardiff University	20	13	6	5	19	18	9	8	1666
The Queen's University of Belfast	0	0	0	0	0	0	0	0	1118

Appendix D9 Inlinks from Different Domains for the UK Biology Departments from AltaVista

University Name	External	com	edu	uk	org	net	co.uk	ac.uk	Web pages
University of Bath	73	10	10	32	4	1	3	27	205
Birkbeck College	5666	874	930	1012	361	137	192	776	7755
University of Birmingham	381	68	26	238	17	14	21	202	348
University of Bristol	1681	607	167	410	102	40	45	332	1709
Brunel University	40	4	7	15	0	0	1	11	660
University of Cambridge	5301	1033	601	1462	625	137	545	785	4692
University of Central Lancashire	82	66	0	12	0	1	0	12	136
Cranfield University	197	40	8	39	18	6	11	24	543
University of Durham	26	3	1	11	3	0	1	10	959
University of East Anglia	115	10	13	70	2	0	7	44	385
University of East London	5	0	0	3	1	0	1	2	123
University of Essex	196	68	5	43	19	6	11	28	561
University of Exeter	61	4	7	16	6	4	1	10	174
University of Hull	60	11	6	23	4	1	1	12	72
Imperial College of Science, Technology and Medicine	1511	202	278	391	144	35	27	321	1737
Keele University	380	47	25	147	19	4	9	72	374
University of Kent at Canterbury	183	22	19	48	8	2	6	37	291
King's College London	664	107	33	285	35	28	33	196	1424
Lancaster University	219	38	8	114	13	5	8	102	228
University of Leeds	582	128	49	194	36	20	22	149	688
University of Leicester	742	168	74	186	71	22	24	135	725
University of Luton	10	0	0	10	0	0	3	6	12
University of Manchester	1455	195	164	779	71	36	78	687	870
University of Manchester Institute of Science & Technology	637	95	131	181	35	14	20	155	1644

Manchester Metropolitan University	44	12	0	31	1	0	2	28	184
University of Newcastle	66	1	0	51	4	0	5	46	218
University of Northumbria at Newcastle	20	4	1	13	0	0	1	10	1
University of Nottingham	11	1	0	9	0	0	2	7	2
Open University	67	13	2	25	7	3	6	12	103
University of Oxford	3023	471	369	874	294	96	130	670	2807
Oxford Brookes University	195	37	12	60	27	1	7	37	703
University of Plymouth	87	5	17	36	2	0	1	35	148
Queen Mary, University of London	190	9	10	132	12	1	3	125	35
University of Reading	572	21	21	409	15	9	24	370	1221
Royal Holloway, University of London	134	43	12	32	13	4	4	22	103
University of Sheffield	234	38	28	61	18	12	4	44	605
University of Southampton	776	46	9	636	22	5	38	585	818
University of Sussex	2175	600	331	357	167	129	101	223	4005
University College London	14470	662	2623	2033	433	442	169	1834	110744
University of Warwick	641	62	97	182	62	13	19	152	674
University of Westminster	14	2	1	8	1	0	2	5	108
University of York	357	61	56	127	29	9	14	96	435
University of Aberdeen	4	0	2	0	0	0	0	0	6
University of Dundee	374	68	30	169	22	3	21	135	805
University of Edinburgh	693	91	41	386	35	9	19	347	1107
University of Glasgow	777	125	122	181	145	26	22	120	3503
University of St Andrews	202	14	6	151	9	3	5	145	319
University of Stirling	449	427	0	20	0	0	2	18	196
University of Wales, Bangor	154	42	9	74	7	2	21	46	789
University of Wales, Swansea	270	39	5	120	22	14	26	63	124

University of Wales, Aberystwyth	266	87	11	121	9	12	11	39	299
Cardiff University	714	172	108	222	72	25	38	133	1068
The Queen's University of Belfast	205	34	36	28	12	7	5	15	1138

Appendix E Correlation Coefficient Values between Average RAE Scores and Research Incomes for the UK Departments and Institutions

Appendix E1 Pearson Correlation Coefficient Values

Units	Estimated research productivities versus research income	RAE averages versus research income average	Numbers
Biology departments	0.951**	0.777**	53
Chemistry departments	0.946**	0.829**	41
Physics departments	0.872**	0.432**	44
All institutions	0.951**	0.666**	173

Appendix E2 Spearman Correlation Coefficient Values

Units	Estimated research productivities versus research income	RAE averages versus research income average	Numbers
Biology departments	0.938**	0.812**	53
Chemistry departments	0.955**	0.839**	41
Physics departments	0.882**	0.457**	44
All institutions	0.932**	0.772**	173

Notes: Estimated research productivities = RAE averages * academic staff members
** = significant at the 1% level

Appendix F Top Level Domains Departments Outlink to

Appendix F1 Some Top Level Domains the UK Physics, Chemistry and Biology Departments Outlink to

Physics			Chemistry			Biology		
domain	links	proportion	Domain	links	proportion	domain	links	proportion
edu	26984	0.204	Gov	19319	0.167	uk	36426	0.439
org	23934	0.181	Edu	18241	0.158	com	9531	0.115
uk	21269	0.161	Uk	16005	0.138	org	8196	0.099
com	19796	0.150	Org	14469	0.125	edu	7073	0.085
gov	6555	0.050	Com	13884	0.120	gov	6206	0.075
fr	5557	0.042	De	10092	0.087	ch	4175	0.050
de	4374	0.033	Jp	8039	0.069	de	2716	0.033
ch	2764	0.021	Au	1589	0.014	fr	1590	0.019
it	2356	0.018	Nl	1507	0.013	jp	1066	0.013
net	2329	0.018	Dk	1501	0.013	pl	724	0.009
es	2231	0.017	Ca	1290	0.011	net	674	0.008
ca	1761	0.013	Se	1039	0.009	se	607	0.007
nl	1715	0.013	Ch	1028	0.009	nl	429	0.005
au	1690	0.013	It	976	0.008	au	390	0.005
jp	767	0.006	Fr	801	0.007	il	386	0.005
se	733	0.006	Net	669	0.006	999	358	0.004
dk	459	0.003	Nz	351	0.003	ca	358	0.004

Appendix F2 Some Top Level Domains the Australian Physics, Chemistry and Biology Departments Outlink to

Physics			Chemistry			Biology		
Domain	links	proportion	Domain	links	proportion	domain	links	proportion
Com	18235	0.357	Com	3382	0.400	**au**	**2223**	**0.314**
Au	**8471**	**0.166**	**Au**	**1894**	**0.224**	com	2207	0.312
Edu	7000	0.137	Org	773	0.091	edu	786	0.111
Org	5566	0.109	Edu	762	0.090	org	453	0.064
Uk	**2171**	**0.042**	**Uk**	**420**	**0.050**	**uk**	**245**	**0.035**
Gov	2161	0.042	De	204	0.024	gov	212	0.030
De	948	0.019	Gov	199	0.023	999	197	0.028
Net	888	0.017	Net	134	0.016	net	166	0.023
Fr	693	0.014	Se	105	0.012	nz	112	0.016
Ca	550	**0.011**	Nl	93	0.011	de	94	0.013
It	530	0.010	Tw	87	0.010	**ca**	79	**0.011**
Jp	438	0.009	**Ca**	65	**0.008**	za	58	0.008
Ch	420	0.008	Ch	37	0.004	fr	36	0.005
Nl	398	0.008	At	32	0.004	ch	29	0.004
Se	389	0.008	It	24	0.003	nl	28	0.004
Es	236	0.005	Fr	21	0.002	jp	26	0.004
999	197	0.004	No	20	0.002	at	22	0.003

Appendix F3 Some Top Level Domains the Canadian Physics, Chemistry and Biology Departments Outlink to

Physics			Chemistry			Biology		
Domain	links	proportion	domain	links	proportion	domain	links	proportion
Com	14105	0.253	com	4801	0.286	com	9729	0.250
Edu	12292	0.221	**ca**	**3453**	**0.206**	**ca**	**7557**	**0.194**
Org	7322	0.132	edu	2839	0.169	edu	6185	0.159
Ca	**7318**	**0.132**	org	2137	0.127	org	5550	0.143
Gov	6074	0.109	**uk**	**807**	**0.048**	gov	3813	0.098
Uk	**1844**	**0.033**	gov	542	0.032	**uk**	**1841**	**0.047**
De	967	0.017	de	386	0.023	de	1096	0.028
Net	871	0.016	net	323	0.019	net	549	0.014
Ch	714	0.013	ch	136	0.008	fr	257	0.007
Au	**608**	**0.011**	**au**	**128**	**0.008**	jp	239	0.006
Fr	574	0.010	fr	126	0.008	nl	218	0.006
Nl	394	0.007	nl	125	0.007	**au**	**212**	**0.005**
Se	344	0.006	jp	123	0.007	ch	197	0.005
999	296	0.005	se	105	0.006	se	197	0.005
Jp	239	0.004	999	95	0.006	999	183	0.005
It	174	0.003	it	77	0.005	us	149	0.004
Il	128	0.002	es	40	0.002	dk	97	0.002

Appendix G Results for International Peer Interlinking

Appendix G1 Number of International Inlinks from the Same Type of Departments

Country from to	Physics	Chemistry	Biology
Canada to UK	107	242	236
UK to Canada	241	158	56
Australia to UK	259	109	85
UK to Australia	368	113	41
Australia to Canada	147	5	46
Canada to Australia	72	27	43

Appendix G2 Link Propensities for Each Set of Departments from the Same Type in Different Countries

Country from to	Physics		Chemistry		Biology	
	staff	page	staff	page	staff	page
Canada to UK	6.99e-5	1.57e-8	2.00e-4	2.09e-7	3.45e-5	4.01e-8
UK to Canada	1.57e-4	3.55e-8	1.31e-4	1.36e-7	8.11e-6	9.43e-9
Australia to UK	2.49e-4	5.15e-8	1.40e-4	2.61e-7	2.77e-5	3.9e-8
UK to Australia	3.53e-4	7.32e-8	1.45e-4	2.7e-7	1.34e-5	1.93e-8
Australia to Canada	2.87e-4	5.79e-8	9.51e-6	2.14e-8	1.51e-5	8.94e-8
Canada to Australia	1.40e-4	2.84e-8	5.13e-5	1.16e-7	1.41e-5	8.36e-8

Appendix H Some Special Target Pages

Appendix H1 Some Special Target Pages for the Australian Physics Departments

Type of target page	URLs
Job information	http://www.physics.adelaide.edu.au/jobs/AIP_JobsReview99.html http://www.physics.adelaide.edu.au/jobs/AusJobs.html http://www.physics.adelaide.edu.au/aip-sa/PhysicsCareers.html http://www.physics.adelaide.edu.au/jobs/Jobs.html
National organizations	http://aos.physics.mq.edu.au/ http://www.physics.adelaide.edu.au/ASGRG/ http://www.physics.adelaide.edu.au/itp/ http://www.physics.adelaide.edu.au/aip-sa/index.htm http://www.phys.utas.edu.au/physics/AIP_TasBranch/ http://www.physics.usyd.edu.au/~obyrne/physics.html http://www.physics.usyd.edu.au/~merchant/WebRing/index.html http://www.physics.usyd.edu.au/aipaust/ http://www.ph.unimelb.edu.au/epp/aushep/ http://astro.ph.unimelb.edu.au/~rwebster/vo_index.html http://msowww.anu.edu.au/~jrm/NCA/ http://www.mso.anu.edu.au/~bessell/DOrep.html http://www.mso.anu.edu.au/~bessell/DOapp.html
Student society	http://ugrad.phys.unsw.edu.au/physoc/index.htm
Information for attending a conference	http://www.mso.anu.edu.au/~iauxxv/visa_information.shtml http://msowww.anu.edu.au/~stanford/funding.html

Appendix H2 Some Special Target Pages for the Canadian Physics Departments

Types of target pages	URLs
Jobs information	http://www.astro.utoronto.ca/~hall/jobs.html http://www.physics.mcgill.ca/pos/astro-03.html
International organization	http://www.physics.umanitoba.ca/IUPAP/IUPAP.html
Student society	http://physics.usask.ca/~pssweb/ http://www.physics.ubc.ca/~physsoc/ http://physics.usask.ca/~pssweb/ http://www.ugrad.physics.mcgill.ca/~msps/explorer.html
Journal	http://pasp.phys.uvic.ca/ http://www.astro.ubc.ca/E-Cass/
Paper	http://www.astro.ubc.ca/E-Cass/VE-98/matthews/index.html
Thesis	http://csr.phys.ualberta.ca/~macqueen/thesis/thesis.html

Appendix H3 Some Special Target Pages for the Canadian Chemistry Departments

Types of target page	URLs
Jobs information	http://www.usask.ca/chemistry/employment.html
National organizations	http://www.chem.ucalgary.ca/csc2000/ http://www.chem.ucalgary.ca/groups/csc-dic/inorgdiv.html http://www.kingsu.ab.ca/~chemed/welcome.htm
Journal	http://www.uwo.ca/chem/canjchem/canadianjournal/

Appendix H4 Some Special Target Pages for the Canadian Biology Departments

Type of target pages	URLs
National organizations	http://www.uoguelph.ca/botany/cba/index.htm http://www.uoguelph.ca/botany/cba/cba.htm http://www.zoology.ubc.ca/~otto/CSEE.html http://www.trentu.ca/biology/botany/welcome.htm http://www.botany.ubc.ca/cccm/
Paper	http://ww2.mcgill.ca/biology/faculty/hendry/HendryEvolution2002.pdf
Book	http://www2.biology.ualberta.ca/palmer/thh/crayfish.htm

Appendix H5 Some Special Target Pages for the UK Physics Departments

Type of target pages	URLs
Jobs information	http://www.hep.phys.soton.ac.uk/~evans/rumour.html http://www.ph.ed.ac.uk/admin/vacancies/vacancies.html http://www.physics.gla.ac.uk/ppt/JOB02/ http://www.physics.gla.ac.uk/ppt/ppt_post.html http://www.hep.phys.soton.ac.uk/~evans/rumour.html
Paper	http://www.mrao.cam.ac.uk/ppeuc/astronomy/papers/baker/baker.html
International or national organizations	http://www.cm.ph.bham.ac.uk/uksf/main http://www-astro.physics.ox.ac.uk/cmbnet/

Appendix H6 Some Special Target Pages for the UK Chemistry Departments

Types of target pages	URLs
Guidance to find the department	http://www.chem.bham.ac.uk/findus.htm
	http://www.chem.ucl.ac.uk/resources/bigmap.htm
	http://www.chem.ucl.ac.uk/topimages/map.gif
	http://www.bangor.ac.uk/ch/find_us.html
Sever statistics	http://www.ch.ic.ac.uk/usage/list.html
	http://www.chem.rdg.ac.uk/usage/index.html
	http://www2.shef.ac.uk/chemistry/stats/
Others	http://www.ch.ic.ac.uk/heritage/
	http://www.chm.bris.ac.uk/admin/phonenum.htm
International or national Chemistry organizations	http://www.uea.ac.uk/che/UKCCN/
	http://www.chem.gla.ac.uk/~bob/ukmac.html
	http://www.chem.qmul.ac.uk/iubmb/
Student Society	http://www.chem.leeds.ac.uk/chem-soc/
	http://ch-www.st-and.ac.uk/chemsoc/
Online journals	http://www.ch.cam.ac.uk/ChemJournals.html
Papers	http://www.chem.leeds.ac.uk/papers/html/Nimes/nimes.html
	http://www2.shef.ac.uk/chemistry/www-publications/4_02963A.html
	http://www.ch.ic.ac.uk/ectoc/ectoc-3/pub/008/index.htm
	http://www.ch.ic.ac.uk/ectoc/ectoc-3/pub/033/index.htm
	http://www.ch.ic.ac.uk/ectoc/ectoc-3/pub/005/index.htm
	http://www.ch.ic.ac.uk/ectoc/ectoc-3/pub/042/
	http://www.ch.ic.ac.uk/ectoc/echet98/pub/048/
	http://www.ch.ic.ac.uk/ectoc/papers/49/

Appendix H7 Some Special Target Pages for the UK Biology Departments

Types of target pages	URLs
Professor's CV	http://www-cryst.bioc.cam.ac.uk/~tom/cv.html
Student's homepage	http://student.cryst.bbk.ac.uk/~esodh01/
Jobs information	http://www.bioch.ox.ac.uk/aspsite/index.asp?sectionid=vacancies
National organizations	http://www.kcl.ac.uk/kis/schools/life_sciences/biomed/bscb/top.html http://www.cryst.bbk.ac.uk/BBS/index.html http://bca.cryst.bbk.ac.uk/bca/welcome.htm http://www.dundee.ac.uk/lifesciences/bsdb/
Databases	http://www-cryst.bioc.cam.ac.uk/data/align/ http://www-cryst.bioc.cam.ac.uk/~campass/ http://www-cryst.bioc.cam.ac.uk/~ddbase/ http://www.biochem.ucl.ac.uk/bsm/cath/ http://www.biochem.ucl.ac.uk/bsm/cath_new/cath_info.html http://www.biochem.ucl.ac.uk/bsm/enzymes/index.html http://www.biochem.ucl.ac.uk/bsm/pdbsum/index.html http://www.biochem.ucl.ac.uk/bsm/dbbrowser/jj/owlinfo.html http://www.bioinf.man.ac.uk/dbbrowser/OWL/ http://www.biochem.ucl.ac.uk/bsm/PROCAT/PROCAT.html http://www.biology.leeds.ac.uk/staff/dawa/bats/Index.htm http://cpbnts1.bio.ic.ac.uk/gpdd/newselector.asp http://www-cryst.bioc.cam.ac.uk/~homstrad/ http://www.cryst.bbk.ac.uk/peptaibol/home.shtml
Paper	http://www.biology.leeds.ac.uk/staff/tbt/Papers/TT_TREE00.pdf

Appendix I Translation of ISI Subject Categories into RFCD Codes for Chemistry, Physics and Biology by Linda Butler from REPP

PHYSICAL, CHEMICAL AND BIOLOGICAL SCIENCES

240000 PHYSICAL SCIENCES

240100 Astronomical sciences

- Astronomy & Astrophysics

240200 Theoretical and condensed matter physics

- Physics, condensed matter

240300 Atomic and molecular physics; Nuclear and particle physics; Plasma physics

- Physics, atomic, molecular & chemical
- Physics, nuclear
- Physics, particles & fields

240400 Optical physics

- Optics

240500 Classical physics

- No corresponding ISI category

249900 Other physical sciences

- Biophysics
- Instruments & instrumentation
- Microscopy

24X000 General physical sciences

- Physics, multidisciplinary
- Physics, applied
- Physics, fluids & plasmas
- Nuclear science & technology

250000 CHEMICAL SCIENCES		
250100 Physical chemistry (incl. structural)		
	Chemistry, physical	Electrochemistry
250200 Inorganic chemistry		
	Chemistry, inorganic & nuclear	Crystallography
250300 Organic chemistry		
	Chemistry, organic	
250400 Analytical chemistry		
	Chemistry, analytical	Spectroscopy
250500 Macromolecular chemistry		
	Polymer science	
250600 Theoretical and computational chemistry		
	No corresponding ISI category	
259900 Other chemical sciences		
	Chemistry, applied	
25X000 General chemical sciences		
	Chemistry, multidisciplinary	
270000 BIOLOGICAL SCIENCES		
270100 Biochemistry and cell biology		
	Biochemistry and molecular biology	Cell biology
270200 Genetics		
	Genetics & heredity	

270300 Microbiology

Microbiology	

270400 Botany

Plant sciences	

270500 Zoology

Entomology	Ornithology
Zoology	Zoology (ss)

270600 Physiology

No corresponding ISI category	

270700 Ecology and evolution

Biodiversity conservation	Ecology
Evolutionary biology	

270800 Biotechnology

Biotechnology and applied microbiology	Biochemical research methods

279900 Other biological sciences

Biology, miscellaneous	

27X000 General biological sciences

Biology	

Notes from Linda:

(1) The numerical code is the standard Research Fields, Disciplines and Courses (RFCD) classification scheme developed by the Australian Bureau of Statistics.
(2) The names of the numeric codes are in italics
(3) The names of the ISI subject category journal sets that are used to create a journal set for each RFCD sub-field are shown in normal font below the heading.

The RFCD scheme is used throughout the Australian research and higher education systems, and journal sets in Australia have been aligned with this scheme. There is generally little problem from doing this, except in the cross-over between plant/animal/human research. The RFCD scheme distinguishes between these three, while ISI does not always do so. If a particular ISI journal set appeared to predominantly focus on one aspect, it was classified to the relevant RFCD code. If it clearly spanned two, or all three aspects, it was classified to one of the dummy interdisciplinary codes e.g. Virology

Appendix J Postcodes for Departments

Appendix J1 UK Physics Departments' Postcodes

university of bath|bath|sn3 3jr ba2 7ay|
university of birmingham|birmingham|b15 2tt|
university of bristol|bristol|bs8 1th bs8 1tl|
university of cambridge|cambridge|cb2 1tn cb3 0he cb5 8bl cb3 0ha|
university of central lancashire|lancashire|pr1 2he|
university of durham|durham|dh1 3hp dh1 3le |
university of exeter|exeter|ex4 4qj ex4 4ql|
university of hertfordshire|hatfield|al10 9ab|
imperial college london|london|sw7 2az|
keele university|keele|st5 5bg|
university of kent|kent|ct2 7nz ct2 7nr|
king's college london|london|wc2r 2ls|
lancaster university|lancaster|la1 4yw la1 4yb|
university of leeds|leeds|ls2 9jt|
university of leicester|leicester|le1 7rh|
university of liverpool|liverpool|l69 3bx l69 7ze |
liverpool john moores university|birkenhead|l3 5ux ch41 1ld|
loughborough university|loughborough|le11 3tu|
university of manchester|manchester|m13 9pl|
umist|manchester|m60 1qd|
university of newcastle upon tyne|newcastle upon tyne|ne1 7ru|
university of nottingham|nottingham|ng7 2rd|
open university|milton keynes|mk7 6aa|
university of oxford|oxford|ox1 2jd ox1 3pu|

queen mary university of london|london|e1 4ns|
university of reading|reading|rg6 6ah rg6 6af|
royal holloway university of london|surrey|tw20 0ex|
university of sheffield|sheffield|s10 2tn s3 7rh|
university of southampton|southampton|so17 1bj|
university of surrey|surrey|gu2 7xh|
university of sussex|brighton|bn1 9rh bn1 9qh |
university college london|london|wc1e 6bt|
university of warwick|coventry|cv4 7al|
university of york|york|yo10 5dd |
university of edinburgh|edinburgh|eh8 9ju eh8 9yl eh9 3jz |
university of glasgow|glasgow|g12 8qq|
heriot-watt university|edinburgh|eh14 4as|
university of paisley|paisley|pa1 2be|
university of st andrews|st andrews|ky16 9aj ky16 9ss|
university of strathclyde|glasgow|g1 1xq g4 0ng|
university of wales swansea|swansea|sa2 8pp|
university of wales aberystwyth|aberystwyth|sy23 2ax sy23 3bz |
cardiff university|cardiff|cf10 3xq cf24 3yb|
queen's university belfast|belfast|bt7 1nn|

Appendix J2 UK Chemistry Departments' Postcodes

university of bath|bath|ba2 7ay|
university of birmingham|birmingham|b15 2tt|
university of bristol|bristol|bs8 1th bs8 1ts|
university of cambridge|cambridge|cb2 1tn cb2 1ew|
university of durham|durham|dh1 3hp dh1 3le|
university of east anglia|norwich|nr4 7tj|
university of exeter|exeter|ex4 4qj ex4 4qd|
university of huddersfield|huddersfield|hd1 3dh|
university of hull|hull|hu6 7rx|
imperial college london|london|sw7 2az|
keele university|keele|st5 5bg|
king's college london|london|wc2r 2ls|
university of leeds|leeds|ls2 9jt|
loughborough university|leicestershire|le11 3tu|
university of manchester|manchester|m13 9pl|
umist|manchester|m60 1qd|
university of newcastle upon tyne|newcastle upon tyne |ne1 7ru|
northumbria university |newcastle upon tyne|ne1 8st|
university of nottingham|nottingham|ng7 2rd|
nottingham trent university|nottingham|ng1 4bu ng11 8ns |
open university|milton keynes|mk7 6aa|
university of oxford|oxford|ox1 2jd ox1 3qr ox1 3qz ox1 3ta ox1 3qu|
queen mary university of london|london|e1 4ns|
university of reading|reading|rg6 6ah rg6 6ad|
university of sheffield|sheffield|s10 2tn s3 7hf|
university of southampton|southampton|so17 1bj|

university of surrey|surrey|gu2 7xh|
university of sussex|brighton|bn1 9rh bn1 9qj |
university college london|london|wc1e 6bt wc1h 0aj |
university of warwick|coventry|cv4 7al|
university of york|york|yo10 5dd |
university of aberdeen|aberdeen|ab24 3fx ab24 3ue |
university of edinburgh|edinburgh|eh8 9yl eh9 3jj |
university of glasgow|glasgow|g12 8qq|
heriot-watt university|edinburgh|eh14 4as|
university of st andrews|st andrews|ky16 9aj ky16 9st |
university of strathclyde|glasgow|g1 1xq g1 1xl|
university of wales bangor|bangor|ll57 2dg ll57 2uw|
university of wales swansea|swansea|sa2 8pp|
cardiff university|wales cardiff|cf10 3xq cf10 3at cf10 3tb|
queen's university belfast |belfast |bt7 1nn|

Appendix J3 UK Biology Departments' Postcodes

university of bath|bath|ba2 7ay|
birkbeck university of london|london|wc1e 7hx|
university of birmingham|birmingham|b15 2tt|
university of bristol|bristol|bs8 1th bs8 1ug bs8 1td|
brunel university|middlesex|ub8 3ph|
university of cambridge|cambridge|cb2 1tn cb2 1ga cb2 1qw cb2 3eh cb2 3ea cb2 3ej cb4 0we cb2 1qt cb2 1rx |
university of central lancashire|lancashire|pr1 2he|
cranfield university|cranfield|mk43 0al mk45 4dt |
university of durham|durham|dh1 3hp dh1 3le |
university of east anglia|norwich|nr4 7tj|
university of east london||e15 4lz|
university of essex|colchester|co4 3sq|
university of exeter|exeter|ex4 4qj ex4 4ps|
university of hull|hull|hu6 7rx|
imperial college london|london|sw7 2az sw7 2ay sw7 2bw sl5 7py |
keele university|keele|st5 5bg|
university of kent|kent|ct2 7nz ct2 7nj |
king's college london|london|wc2r 2ls|
lancaster university|lancaster|la1 4yw la1 4yq |
university of leeds|leeds|ls2 9jt|
university of leicester|leicester|le1 7rh|
university of luton|luton|lu1 3ju lu1 5du |
university of manchester|manchester|m13 9pl m13 9pt|
umist|manchester|m60 1qd|
manchester metropolitan university|manchester|m15 6bh m1 5gd|
university of newcastle upon tyne|newcastle upon tyne |ne1 7ru ne2 4hh ne1 3bz |

northumbria university |newcastle upon tyne|ne1 8st|
university of nottingham|nottingham|ng7 2rd|
open university|milton keynes|mk7 6aa|
university of oxford|oxford|ox1 2jd ox1 3rb ox1 3ps ox1 3qu|
oxford brookes university|oxford|ox3 0bp|
university of plymouth|plymouth|pl4 8aa|
queen mary university of london |london|e1 4ns|
university of reading|reading|rg6 6ah rg6 6aj|
royal holloway university of london|surrey|tw20 0ex|
university of sheffield|sheffield|s10 2tn|
university of southampton|southampton|so17 1bj so16 7px|
university of sussex|brighton|bn1 9rh bn1 9qg|
university college london|london|wc1e 6bt|
university of warwick|coventry|cv4 7al|
university of westminster|london|w1b 2uw w1w 6uw|
university of york|york|yo10 5dd yo10 5yw |
university of aberdeen|aberdeen|ab24 3fx ab24 2tz|
university of dundee|dundee|dd1 4hn dd1 5eh|
university of edinburgh|edinburgh|eh8 9yl eh9 3jr eh9 3jt eh9 3jq|
university of glasgow|glasgow|g12 8qq|
university of st andrews|st andrews|ky16 9ts ky16 9aj|
university of stirling|stirling |fk9 4la|
university of wales bangor|bangor|ll57 2dg ll57 2uw|
university of wales swansea|swansea|sa2 8pp|
university of wales aberystwyth|aberystwyth|sy23 2ax sy23 3da sy23 3dd |
cardiff university|wales cardiff|cf10 3xq cf10 3us|
queen's university belfast |belfast |bt7 1nn bt9 7bl|

Appendix J4 Canadian Physics Departments' Postcodes

university of alberta|edmonton|t6g 2e1 t6g 2j1|
brandon university|manitoba|r7a 6a9|
university of british columbia|vancouver|v6t 1z4 v6t 1z1|
brock university|st. catharines|l2s 3a1|
university of calgary|calgary|t2n 1n4|
carleton university|ottawa|k1s 5b6|
university of concordia|montreal|h3g 1m8 h4b 1r6|
university of dalhousie|halifax|b3h 4r2 b3h 3j5 |
université de moncton|edmundston moncton|e3v 2s8 e1a 3e9|
université de sherbrooke|sherbrooke |j1k 2r1|
école polytechnique|montreal|h3c 3a7|
universtiy of guelph|guelph|n1g 2w1 |
institut national de recherche scientifique|varennes|j3x 1s2|
lakehead university|thunder bay|p7b 5e1|
université laval||g1k 7p4|
university of manitoba|winnipeg|r3t 2n2|
mcgill university|montreal|h3a 2t5 h3a 2t8 |
mcmaster university|hamilton|l8s 4l8 l8s 4m1|
memorial university of newfoundland|st. john's|a1c 5s7 a1b 3x7|
university of montréal|montreal|h3c 3j7 |
mount allison university|sackville|e4l 1e6|
university of new brunswick|fredericton saint john|e3b 5a3 e2l 4l5|
university of ottawa|ottawa|k1n 6n5|
university of prince edward island|charlottetown|c1a 4p3|
québec à trois-rivières|trois-rivieres|g9a 5h7|
queen's university at kingston|kingston|k7l 3n6|
university of regina|regina|s4s 0a2|

ryerson university|toronto|m5b 2k3|
university of saint mary's|halifax|b3h 3c3|
university of saskatchewan|saskatoon|s7n 5e2|
simon fraser university|burnaby|v5a 1s6|
university of toronto|toronto|m5s 1a1 m5s 1a7 |
university of victoria|victoria|v8w 2y2 v8w 3p6 v8p 5c2|
university of waterloo|waterloo|n2l 3g1|
university of western ontario|london|n6a 5b8 n6a 3k7|
university of winnipeg|winnipeg|r3b 2e9|
york university|toronto|m3j 1p3|

Appendix J5 Canadian Chemistry Departments' Postcodes

acadia university|wolfville|b4p 2r6 b0p 1x0|
university of alberta|edmonton|t6g 2e1 t6g 2g2|
university of british columbia|vancouver|v6t 1z4 v6t 1z1|
university of brock|st. catharines|l2s 3a1|
university of calgary|calgary|t2n 1n4|
carleton university|ottawa|k1s 5b6 |
concordia university|montreal|h3g 1m8 h4b 1r6|
dalhousie university|halifax|b3h 3j5 b3h 4j3|
université de sherbrooke|sherbrooke|j1k 2r1|
québec à montréal|montreal|h3c 3p8|
university of guelph|guelph|n1g 2w1 |
the king's university college|edmonton|t6b 2h3|
lakehead university|thunder bay|p7b 5e1|
université laval|quebec|g1k 7p4|
university of lethbridge|lethbridge|t1k 3m4 |
university of manitoba|winnipeg|r3t 2n2|
mcgill university|montreal|h3a 2t5 h3a 2k6 |
mcmaster university|hamilton|l8s 4l8 l8s 4m1|
university of montréal|montreal|h3c 3j7 h3t 1j4|
mount allison university|sackville|e4l 1g7|
university of new brunswick|fredericton|e3b 5a3 e3b 6e2|
university of northern british columbia|prince george|v2n 4z9|
university of ottawa|ottawa|k1n 6n5|
university of prince edward island|charlottetown|c1a 4p3|
queen's university|kingston|k7l 3n6|
québec à rimouski|rimouski |g5l 3a1 g6v 8r9|
university of regina|regina|s4s 0a2|

ryerson university|toronto|m5b 2k3|
saint mary's university|halifax|b3h 3c3|
university of saskatchewan|saskatoon|s7n 5c8 s7n 5c9|
simon fraser university|burnaby|v5a 1s6|
st. francis xavier university|antigonish|b2g 2w5|
university of toronto|toronto|m5s 1a1 m5s 3h6 m5s 3e5|
trent university|peterborough|k9j 7b8|
university of victoria|victoria|v8w 2y2 v8p 5c2 v8w 3v6|
university of waterloo|waterloo|n2l 3g1|
university of western ontario|london|n6a 5b8 n6a 5b7|
wilfrid laurier university|waterloo|n2l 3c5 n2l 3g1|
university of winnipeg|winnipeg|r3b 2e9|
york university|toronto|m3j 1p3|

Appendix J6 Canadian Biology Departments' Postcodes

acadia university|wolfville|b4p 2r6 b0p 1x0|
university of alberta|edmonton|t6g 2e1 t6g 2e9 t6g 2p5|
brandon university|brandon|r7a 6a9|
university of british columbia|vancouver|v6t 1z4 v6t 1z3 |
university of brock|st. catharines|l2s 3a1|
university of calgary|calgary|t2n 1n4 t2n 4n1|
concordia university|montreal|h3g 1m8 h4b 1r6|
dalhousie university|halifax|b3h 3j5 b3h 1x5 b3h 4j1|
university of guelph|guelph|n1g 2w1|
lakehead university|thunder bay|p7b 5e1|
laurentian university|sudbury|p3e 2c6|
université laval|quebec|g1k 7p4|
university of lethbridge|lethbridge|t1k 3m4|
university of manitoba|winnipeg|r3t 2n2 r3e 3j7|
mcgill university|montreal|h3a 2t5 h3g 1y6 h9x 3v9 h3a 2b4 h3a 1b1 h3a 2b2|
mcmaster university|hamilton| l8s 4l8 l8s 4k1|
memorial university of newfoundland|st. john's|a1c 5s7 a1b 3x9|
university of montréal|montreal|h3c 3j7 j2s 2m2 j2s 7c6|
mount allison university|sackville|e4l 1g7|
university of new brunswick|fredericton saint john|e3b 5a3 e2l 4l5 e3b 6e1|
university of northern british columbia|prince george|v2n 4z9|
nova scotia agricultral college|truro|b2n 5e3|
okanagan university college|kelowna|v1v 1v7|
university of ottawa|ottawa|k1n 6n5 klh 8m5 |
québec à montréal|montreal|h3c 3p8 h3b 3h5|
québec à rimouski|rimouski|g5l 3a1 g6v 8r9|
queen's university|kingston|k7l 3n6|

university of regina|regina|s4s 0a2|
ryerson university|toronto|m5b 2k3|
saint mary's university|halifax|b3h 3c3|
unviersity of saskatchewan|saskatoon|s7n 5c8 s7n 5b4 s7n 5e5 s7n 5e2 s7n 5a8|
université de sherbrooke|sherbrooke |j1k 2r1 j1h 5n4|
simon fraser university|burnaby|v5a 1s6|
st. francis xavier university|antigonish|b2g 2w5 |
university of toronto|toronto|m5s 1a1 m5s 1a8 m5s 3g5 m1c 1a4 m5s 3e2|
trent university|peterborough|k9j 7b8|
university of victoria|victoria| v8w 2y2 v8p 5c2 v8w 3n5 v8w 3p6 |
university of waterloo|waterloo|n2l 3g1|
university of western ontario|london|n6a 5b8 n6a 5b7 n6a 5c1|
york university|toronto|m3j 1p3|